Journal of Theoretical and Applied Information Technology
31st July 2025. Vol.103. No.14
© Little Lion Scientific

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CROSS-MODAL ADAPTIVE META-FREE LEARNING FOR SCALABLE CONTINUAL ZERO-SHOT GENERALIZATION

RAFIAH S B[1], B PRAJNA[2]

[1] Research Scholar, Department of Computer Science and Systems Engineering, Andhra University, Visakhapatnam, India.

[2]Professor, Department of Computer Science and Systems Engineering, Andhra University, Visakhapatnam, India.

Email: [1]rafiah.sb@gmail.com [2]prajna.mail@gmail.com

ABSTRACT

Continual Zero-Shot Learning (CZSL) involves training models to learn sequentially from separate data streams while effectively generalizing to unseen classes without revisiting earlier data. However, conventional approaches often face issues such as catastrophic forgetting and weak generalization, especially under noisy, multimodal, or low-resource conditions. To overcome these limitations, the proposed work introduced Cross-Modal Adaptive Meta-Free Learning (CAMeL) a scalable, task-free learning framework. CAMeL incorporated a Cross-Modal Generative Memory to synthesize both visual and semantic features, ensuring knowledge retention across tasks. It also features a Neural Attribute Synthesizer that generates context-aware prompts, enhancing adaptability to challenging learning conditions. The framework is further optimized through Continual Learning Adaptive Sharpness-Aware Minimization (CLASAM), which flattens the loss landscape to promote stability and generalization. CAMeL effectively supports multimodal learning, reduces forgetting, and handles both zero-shot and few-shot tasks. Experimental results across six benchmarks including CUB, AWA1, and SUN show that CAMeL+CLASAM achieves up to 7.5% higher harmonic mean than existing methods, proving its robustness and scalability.

Keywords: *Continual Learning, Zero-Shot Learning, Sharpness-Aware Optimization, Prompt Learning, Cross-Modal Learning*

1. INTRODUCTION

This study was initiated to address the growing challenge of building scalable and adaptive Artificial Intelligence (AI) systems capable of learning continuously from dynamic environments without forgetting previous knowledge or retraining on past data. The rapid evolution of AI and machine learning has enabled the development of models capable of learning from sequential data streams and generalizing to unseen categories [1], [2]. Two foundational paradigms supporting these advancements are Continual Learning (CL) and Zero-Shot Learning (ZSL) [3]. CL focuses on the progressive acquisition of knowledge over time without suffering from catastrophic forgetting a phenomenon where previously learned information is overwritten [4]. In contrast, ZSL targets the classification of instances from novel classes by exploiting semantic relationships between known and unknown categories [5]. The integration of these approaches has led to the emergence of Continual Zero-Shot Learning (CZSL) [6], which aims to build models that can learn continuously from evolving data streams while still generalizing to new, unseen classes without the need for retraining.

Although notable advancements have been made, existing CZSL methods still encounter several challenges. Most approaches depend on fixed task boundaries; require storing previous data, and use single modality learning models constraints that reduce their effectiveness in real world scenarios involving diverse data types and minimal supervision [7]. Additionally, catastrophic forgetting remains unresolved in scenarios involving noisy inputs and hybrid zero/few-shot conditions [8]. These challenges hinder the scalability, adaptability, and robustness of existing CZSL models.

Despite the integration of CL and ZSL, existing approaches struggle to operate without predefined task boundaries and often depend on stored data or fixed modalities. The core problem lies in catastrophic forgetting and limited generalization to unseen or noisy classes in dynamic environments. This is a major concern because real-world systems such as intelligent surveillance, medical diagnosis, and autonomous vehicles frequently encounter evolving, unseen inputs where retraining is not feasible. Thus, the lack of a scalable, memory-efficient, and adaptive framework limits the practical deployment of continual zero-shot learning models. This study specifically targets these limitations.

To address these challenges, this study proposes a framework that avoids task boundaries, eliminates the need to store past data, and supports multimodal input. CAMeL is designed to synthesize both visual and semantic features, thereby enhancing memory retention without requiring raw data. Furthermore, it dynamically generates prompts based on contextual semantics, which enables robust generalization to novel and noisy inputs. The rationale behind this architecture is its compatibility with real-world conditions where memory buffers are limited, supervision is sparse, and unseen categories are common. CAMeL's optimization through CLASAM ensures model stability by minimizing sharpness in the loss landscape, further boosting its ability to retain prior knowledge while adapting to new distributions. This makes the proposed approach both practically viable and theoretically grounded for scalable continual zero-shot learning.

The originality of this study lies in the integration of cross-modal generative memory and neural prompt synthesis within a task-free continual learning framework. Unlike existing CZSL models, CAMeL does not rely on task boundaries or stored replay data, and it introduces CLASAM to ensure generalization by flattening sharp loss landscapes.

To address these gaps, this paper introduced a novel framework called Cross-Modal Adaptive Meta-Free Learning (CAMeL) [9], aimed at improving scalability and generalization in CZSL. CAMeL features two primary innovations: (i) a Cross-Modal Generative Memory that synthesizes visual and semantic features [10] to enable knowledge retention without relying on explicit data storage, and (ii) a Neural Attribute Synthesizer that produces context-aware prompts to improve generalization in noisy and unseen settings [11]. Additionally, CAMeL is designed to handle multimodal inputs and few-shot scenarios, making it highly adaptable to versatile and practical hybrid learning settings.

The main contributions of this research are summarized as follows:

- This study proposed a task-free, memory-efficient continual zero-shot learning [12] framework that synthesizes cross-modal representations to mitigate catastrophic forgetting.
- The proposed work introduced a dynamic prompt generation mechanism through a Neural Attribute Synthesizer, enhancing robustness to noise and low-resource environments [13].
- The proposed framework integrated few-shot learning capabilities within the continual learning stream, enabling CAMeL to adapt flexibly to hybrid zero/few-shot settings.
- This study conducted comprehensive evaluations of CAMeL on six benchmark datasets CUB, aPY, AWA1, AWA2, SUN [14], and ImageNet 1K showing that it consistently outperforms existing state-of-the-art CZSL methods.

The rest of the paper is structured as follows: Section II provides an overview of related work in continual learning, zero-shot learning, and cross-modal generalization. Section III details the proposed CAMeL framework along with the CLASAM optimization strategy. Section IV discusses the experimental setup, presents results, and offers in-depth analysis. Lastly, Section V concludes the study and suggests potential avenues for future research.

2. LITERATURE REVIEW

The literature review is critical because it identifies the strengths and limitations of existing Continual Learning, Zero-Shot Learning, and Continual Zero-Shot Learning methods. It highlights key challenges such as catastrophic forgetting, task dependency, poor scalability, and limited generalization under noisy conditions that current models fail to address. By clearly revealing these research gaps, the review justifies the need for the proposed CAMeL framework, which introduces cross-modal synthesis and dynamic prompt generation. Moreover, it establishes the novelty, technical rationale, and scientific credibility of this work, ensuring that CAMeL is recognized as a significant and necessary advancement in scalable continual zero-shot learning.

In 2024, Wang et al. [15] highlighted the significance of continual learning in AI systems, emphasizing its role in enabling adaptive knowledge acquisition and updates over time. A key issue identified is catastrophic forgetting, where learning new tasks negatively impacts performance on previously learned ones. The study offers a thorough review of fundamental theories, methodologies, and practical applications. It stresses the importance of balancing stability and adaptability (plasticity) while ensuring task generalization, and presents taxonomy of strategies designed to tackle the core challenges of continual learning.

In 2025, Aslam et al. [16] proposed an innovative continual learning (CL) model aimed at mitigating catastrophic forgetting in rapidly changing domains such as disease outbreak forecasting. The approach integrates Elastic Weight Consolidation and the Fisher Information Matrix (FIM) to retain prior knowledge while adapting to new information. Experimental results on datasets involving influenza, mpox, and measles demonstrate that the model surpasses existing state-of-the-art techniques, achieving high R-squared scores, reducing forgetting by 65%, and enhancing memory stability by 18%. The study underscores the model's effectiveness in capturing and predicting temporal patterns in dynamic health data.

In 2024, Wu et al. [17] introduced ZEST, a zero-shot learning (ZSL) framework aimed at classifying both known and previously unseen IoT devices, particularly in scenarios where device traffic is unavailable during training. The framework features SANE, a self-attention-based network feature extractor that captures latent patterns in IoT traffic, along with a generative model that creates pseudo samples from these features. A supervised classifier is then trained on this synthetic data for effective device identification. Evaluations using real-world IoT traffic show that ZEST achieves notably higher classification accuracy than baseline methods, with SANE outperforming conventional LSTM-based models in feature extraction quality.

In 2024, Lu et al. [18] introduced PAMK, a prototype-augmented multi-teacher knowledge transfer framework designed for continual zero-shot learning (CZSL). The model aims to maintain stability in recognizing previously learned tasks while enhancing adaptability to new ones. Unlike traditional CZSL approaches that risk negative transfer due to an overemphasis on past knowledge, PAMK incorporates two novel components: Prototype Augmented Contrastive Generation (PACG) and Multi-Teacher Knowledge Transfer (MKT). These modules work together to effectively balance retention and generalization in evolving learning scenarios. PACG employs a continual prototype augmentation strategy based on relevance scores to reduce semantic decay and uses a semantic-visual contrastive loss to enhance intra-class compactness. Meanwhile, MKT leverages semantic knowledge from previous tasks to aid new task recognition, mitigating negative transfer. Experimental results show that PAMK significantly outperforms state-of-the-art methods, achieving notable improvements in mean harmonic accuracy on the CUB (3.28%), AWA1 (3.09%), and AWA2 (3.71%) datasets in task-free CZSL settings.

In 2024, Gautam et al. [19] introduced a generative replay-based continual zero-shot learning (GRCZSL) method to classify unseen classes without forgetting past knowledge. Unlike traditional ZSL, which assumes all seen class samples are available, GRCZSL learns from streaming data. It mitigates catastrophic forgetting by replaying synthetic samples generated using a conditional variational autoencoder. The method, designed for a single-head continual learning setup, is evaluated on five benchmark datasets, outperforming baselines and state-of-the-art approaches for real-world applications.

In 2024, Jiang et al. [20] proposed XProDNet, a Cross-modal Prompt-Driven Network aimed at improving image captioning in low-resource settings, particularly within domains such as medical imaging and non-English languages. Unlike traditional approaches that depend on large volumes of labeled data, XProDNet is capable of producing accurate and detailed captions with minimal supervision. The framework was rigorously evaluated across six benchmark datasets, spanning three application domains (standard, medical, and multilingual captioning), four target languages (English, Chinese, German, and French), and two learning paradigms (fully-supervised and few-shot). Results showed that XProDNet consistently outperforms existing state-of-the-art models, demonstrating strong potential for practical, real-world deployment.

3. PROPOSED METHODOLOGY

This study presents a methodology designed to rigorously evaluate the effectiveness of the proposed CAMeL framework in task-free CZSL scenarios. CAMeL was tested across three diverse and challenging benchmark datasets CUB, AWA1, and SUN to assess its robustness under fine-grained, attribute-based, and large-scale scene recognition tasks. The experimental setup simulated a task-free continual learning environment in which disjoint class batches are presented sequentially without access to prior data. During training, visual features and semantic attributes are jointly encoded and processed through CAMeL's two core modules: the Cross-Modal Generative Memory (CMGM) and the Neural Attribute Synthesizer (NAS). A multimodal replay mechanism is employed to mitigate catastrophic forgetting, while dynamically generated prompts enhance the model's ability to generalize to unseen or noisy classes. Additionally, few-shot samples are incorporated in hybrid scenarios to validate the framework's adaptability under limited supervision. Performance is measured using three core metrics mean seen accuracy (mSA), mean unseen accuracy (mUA), and harmonic mean (mH) along with scalability assessments. Comparative evaluations against state-of-the-art CZSL baselines are conducted to demonstrate CAMeL's superiority in knowledge retention, cross-modal generalization, and continual adaptation in both zero-shot and few-shot learning settings.

To enhance the optimization stability and generalization capacity of CAMeL, The proposed work introduced a novel learning strategy named Continual Learning Adaptive Sharpness-Aware Minimization (CLASAM). Unlike traditional optimizers that minimize loss at fixed parameter points, CLASAM minimizes the worst-case (sharpest) loss within a small neighborhood around the model parameters. This approach helps the model converge to flatter minima in the loss landscape, which are known to generalize better across tasks. At each training step, CLASAM perturbs the model weights slightly to estimate the sharpness of the loss surface, and then updates the weights to avoid sharp, unstable regions while still minimizing the core loss. Uniquely, CLASAM adjusts the perturbation scale adaptively based on context assigning greater attention to noisy or unseen data distributions. CLASAM's adaptability makes it highly compatible with CAMeL, which operates in noisy, task-free, and resource-constrained settings. By encouraging flatter optimization landscapes and minimizing overfitting risks, CLASAM strengthens CAMeL's capacity to sustain consistent performance during continual learning in zero-shot scenarios.

This section outlines the experimental methodology used to evaluate the proposed CAMeL framework. It includes the research design, datasets, and core components of the model, training setup and evaluation metrics. The structured approach ensures reproducibility and highlights how CAMeL addresses the challenges of task-free continual zero-shot learning.

3.1 Research Design

The study adopted a task-free continual zero-shot learning setup. CAMeL was evaluated using disjoint class batches that were introduced sequentially without access to previously seen data. This simulated a real-world continual learning scenario. Few-shot settings were also tested using hybrid zero/few-shot configurations.

1. Datasets

The model was trained and evaluated on three widely used benchmark datasets:

1. CUB: Fine-grained bird classification
2. AWA1: Attribute-based animal classification
3. SUN: Scene recognition dataset

These datasets represent a mix of visual complexity and semantic richness to test the robustness of the framework.

2. Model Components

- Cross-Modal Generative Memory (CMGM): Synthesizes visual-semantic representations from latent codes for replay without storing raw data.
- Neural Attribute Synthesizer (NAS): Dynamically generates semantic prompts based on context to guide classification.
- CLASAM Optimizer: Enhances training by minimizing the sharpest local loss and promoting generalization.

3. Training Setup

- Visual features were encoded using a Vision Transformer.
- Semantic features were encoded using BERT embeddings.
- The latent space was constructed using a variational encoder.
- The optimizer (CLASAM) perturbed model parameters to avoid sharp minima.

4. Evaluation Metrics

Three main metrics were used:

- Mean Seen Accuracy (mSA)
- Mean Unseen Accuracy (mUA)

- Harmonic Mean (mH)

3.2 Cross-Modal Adaptive Meta-Free Learning (CAMEL)

The training process of the proposed CAMeL framework is composed of six major stages, each contributing to scalable, task-free continual zero-shot generalization. Figure 1 Diagrammatic Representation of the CAMEL Framework's Architectural Components.

Step 1: Input Encoding

Description: Raw inputs, including images and semantic attributes, are independently encoded into a shared feature space to facilitate multimodal learning.

Let:

- X denote the visual input space (e.g., images),
- A denote the semantic attribute space (e.g., text embeddings),
- $f_v: x \to \mathbb{R}^d$ be the visual encoder (e.g., Vision Transformer),
- $f_a: A \to \mathbb{R}^d$ be the semantic encoder (e.g., BERT).

Given a sample x∈X and attribute vector a∈A, the features are encoded using Equation (1).

$$\upsilon = f_\upsilon(x), \quad a' = f_a(a) \tag{1}$$

These features are concatenated to form a multimodal representation, as defined by Equation (2).

$$h = Concat(\upsilon, a') \tag{2}$$

Step 2: Latent Space Mapping (Variational Encoding)

Description: The concatenated features are mapped into a latent probabilistic space to enable stochastic sampling for generative replay.

The encoder network $q_\emptyset$ maps h into a latent distribution, as described by Equation (3).

$$q_\phi(z/h) = N(\mu(\mathrm{h}), \sigma(h)^2) \tag{3}$$

Where,

- $\mu(h)$ and $\sigma(h)$ are the learned mean and variance vectors.

Sampling z from this distribution is performed using the reparameterization trick, as described in Equation (4).

$$z = \mu(h) + \sigma(h)\Theta \in, \quad \in \sim N(0,1) \tag{4}$$

This latent variable captures the essential joint information for replay synthesis.

Step 3: Cross-Modal Feature Synthesis (Generative Replay)

Description: To prevent catastrophic forgetting, synthetic multimodal features are generated directly from the latent code z.

A decoder network g_θ reconstructs the synthetic features, as defined in Equation (5).

$$\hat{h} = g_\theta(z) \tag{5}$$

These synthesized features $\hat{h}$ approximate the original h and serve as replay data during continual learning without explicitly storing past samples.

Step 4: Neural Attribute Synthesizer (Prompt Generation)

Description: Dynamic semantic prompts are generated based on class attributes and context (such as noise or domain shifts) to guide model adaptation.

The Neural Attribute Synthesizer s_ω maps semantic embeddings and contextual information c into prompt vectors, as described in Equation (6).

$$p = s_\omega(a', c) \tag{6}$$

where:

- a' is the encoded attribute vector,
- c is an optional context vector representing environmental factors.

The prompt p is then fused with the synthesized features, as described in Equation (7).

$$h' = Concat(\hat{h}, p) \tag{7}$$

This enables the model to dynamically adjust its internal representations for better generalization to unseen and noisy classes.

Step 5: Classification and Prediction

Description: The final representation h' is fed into a classifier ψ to predict the class label, considering both seen and unseen classes.

The prediction is made by equation (8),

$$\hat{y} = \arg \max_{y \in y_S \cup y_u} \psi(h', y) \qquad (8)$$

Where,

y_S are seen classes,

y_u are unseen classes.

This ensures that CAMeL can perform both zero-shot and few-shot recognition continually.

Step 6: Training Objective and Optimization

Description: CAMeL is optimized by minimizing a composite loss function composed of three components:

1. *Reconstruction Loss:* Synthetic features are optimized to approximate real features, as shown in Equation (9).

$$L_{rec} = \left\| h - \hat{h} \right\|^2 \qquad (9)$$

2. *KL Divergence Loss*: Equation (10) regularizes the latent space distribution to align with a standard normal prior.

$$L_{KL} = D_{KL}\left(q_\phi(z/h)\right)\left\| N(o, I) \right) \qquad (10)$$

3. *Prompt Consistency Loss*: Equation (11) aligns the synthesized prompts with the true semantic attributes.

$$L_{prompt} = \left\| p - f_a(a) \right\|^2 \qquad (11)$$

The total loss is a weighted combination by equation (12),

$$L = L_{rec} + \beta L_{KL} + \gamma L_{prompt} \qquad (12)$$

Where, β and γ are hyperparameters balancing the importance of each component.

The network parameters $\emptyset, \theta, \omega, \psi$ are updated using backpropagation and CLASAM optimization.

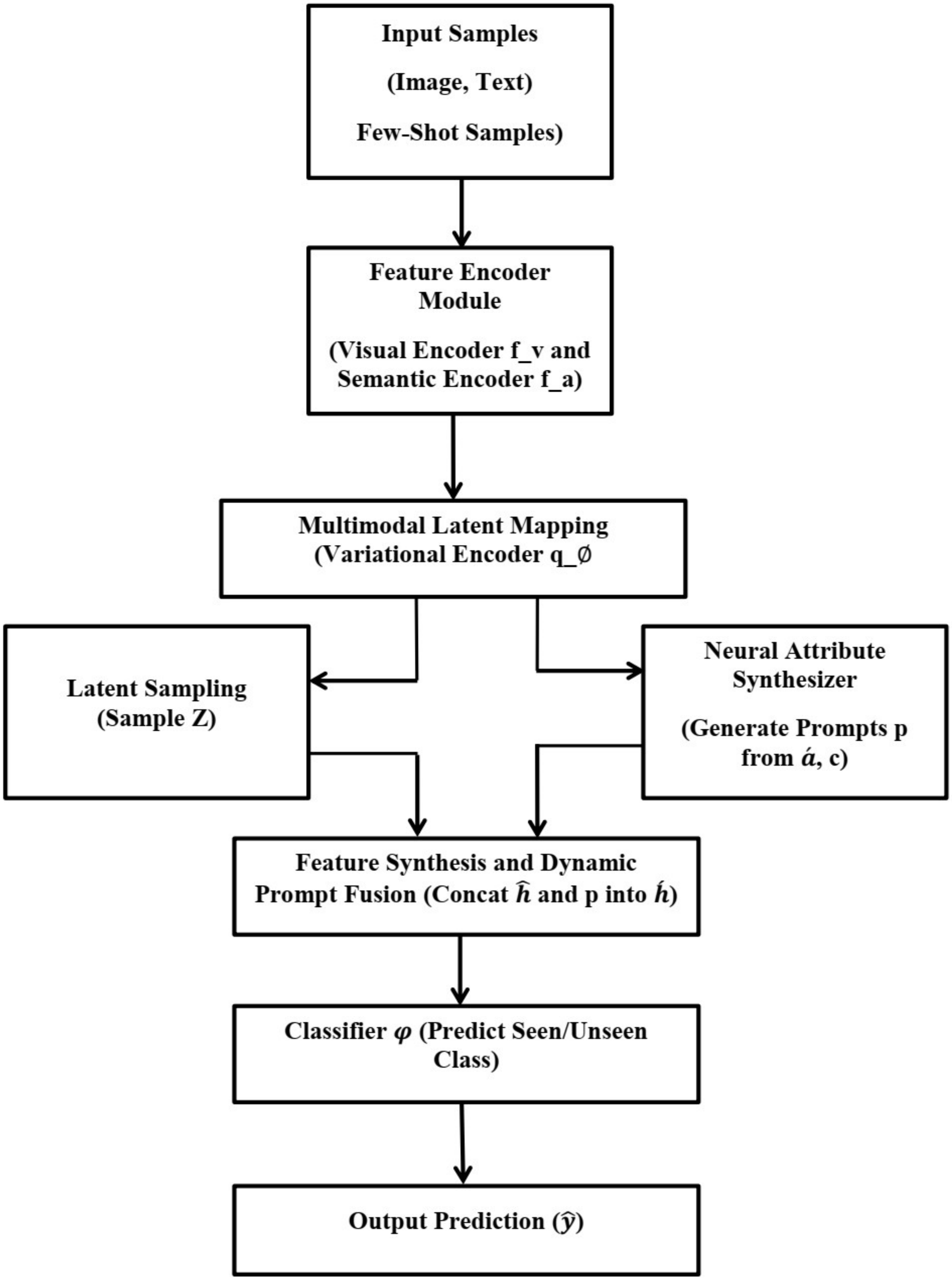
Input Samples
(Image, Text)
Few-Shot Samples)
Feature Encoder Module
(Visual Encoder f_v and Semantic Encoder f_a)
Multimodal Latent Mapping
(Variational Encoder q_∅
Latent Sampling
(Sample Z)
Neural Attribute Synthesizer
(Generate Prompts p from á, c)
Feature Synthesis and Dynamic Prompt Fusion (Concat ĥ and p into h́)
Classifier φ (Predict Seen/Unseen Class)
Output Prediction (ŷ)

Figure 1: Overview of the CAMEL Framework Architecture

3.3 Continual Learning Adaptive Sharpness-Aware Minimization

CLASAM is a novel optimization strategy designed to improve model stability and generalization in continual learning environments. Unlike conventional optimizers that minimize the loss at fixed parameter values, CLASAM seeks to minimize the sharpest possible loss within a small neighborhood around the parameters. This leads to flatter minima in the loss landscape, which are known to generalize better across tasks. At each training step, CLASAM perturbs the model weights slightly to estimate local sharpness and then updates the parameters in a way that both reduces loss and avoids high-curvature regions. A major strength of CLASAM is its dynamic adaptability, where perturbation strength is adjusted according to semantic context, allowing for greater responsiveness to unseen or noisy class distributions. This feature makes CLASAM an excellent fit for the CAMeL framework, which functions in task-free, multimodal, and noise-prone environments. By smoothing the learning path, CLASAM supports CAMeL in preserving past knowledge while efficiently adapting to new tasks and classes, making it highly suitable for continual zero-shot learning scenarios. As shown in Figure 2, the workflow of the Continual Learning Adaptive Sharpness-Aware Minimization Optimization process is illustrated.

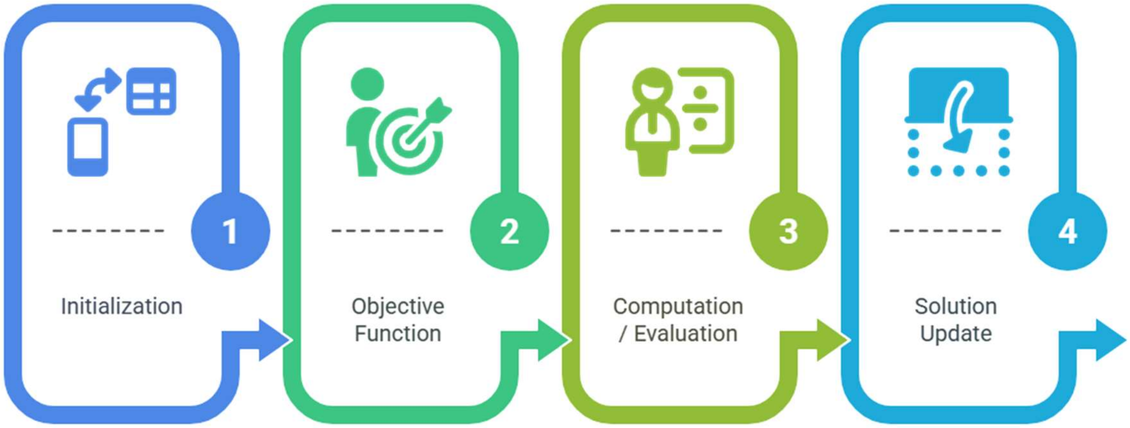

Figure 2: Workflow of the Continual Learning Adaptive Sharpness-Aware Minimization Optimization Process

3.3.1 initialization

In the initialization phase of the CLASAM optimization process, the model begins by setting up its trainable parameters, including the weights of the visual encoder, semantic encoder, latent variable modules, neural attribute synthesizer, and the classifier within the CAMeL framework. Alongside model weights, key optimization hyperparameters are also defined. These include the learning rate η which controls the step size during gradient descent, and the sharpness radius ρ, which determines the scale of perturbation used to assess the sharpness of the loss landscape. Additionally, context-aware scaling factors may be initialized to dynamically adjust ρ, based on data complexity, such as the presence of unseen classes or semantic noise. This initialization step ensures that the CLASAM process begins from a consistent and well-conditioned state, enabling effective training in subsequent stages.

3.3.2 objective function

The core objective of CLASAM is to enhance the stability and generalization capability of the CAMeL framework by optimizing not only the base loss but also the sharpness of the loss landscape. The base objective function in CAMeL is a composite loss that combines three components: reconstruction loss L_{rec} Kullback-Leibler divergence L_{rec} and prompt consistency loss L_{propmt}. These components are weighted by their respective coefficients β and γ, forming the total base loss as defined in Equation (13):

$$L(w) = L_{rec} + \beta L_{KL} + \gamma L_{prompt} \quad (13)$$

CLASAM extends this objective by introducing sharpness-aware minimization, which focuses on minimizing the worst-case loss within a small neighborhood of the current parameters. To achieve this, a perturbation vector ∈ is calculated in the direction of the normalized gradient of the loss. The perturbed objective is then re-evaluated as $L =$

$(w{+}\in)$, which serves as the sharpness-aware loss to be minimized. This dual-objective formulation enables CLASAM to identify parameter updates that not only reduce the primary loss but also encourage convergence to flatter and more generalizable regions in the loss landscape.

3.3.3 computation / evaluation

In the computation and evaluation stage, the model performs a standard forward pass using the current parameters w to compute the base loss *L(w)*. This loss captures the discrepancy between the generated outputs and ground truth values across the reconstruction, KL divergence, and prompts consistency components. Next, the sharpness of the loss landscape is estimated to prepare for sharpness-aware optimization. This is achieved by calculating the gradient $\nabla_w L(w)$, which indicates the direction in which the loss increases most rapidly. This gradient is then normalized and scaled by a sharpness radius ρ to generate a perturbation vector $\in$, such that $\in = \rho . \frac{\nabla_w L(w)}{\|\nabla_w L(w)\|}$. The model parameters are temporarily perturbed to $w' = w{+}\in$, and a second forward pass is conducted to compute the sharpness-aware loss $L(w')$. This perturbed evaluation captures the worst-case behavior of the model within a defined neighborhood and allows the optimizer to identify solutions that are robust to sharp variations in the loss surface.

3.3.4 solution update

In the solution update stage, CLASAM performs backpropagation using the gradient of the sharpness-aware loss $L(w{+}\in)$, rather than the original base loss. This ensures that parameter updates are influenced not just by the immediate loss values but also by the local geometry of the loss surface. The gradient $\nabla_w L(w{+}\in)$, is computed and used to update the model weights, as specified in Equation (14).

$$w \leftarrow w - \eta \ \nabla_w \ L(w+\varepsilon), \qquad (14)$$

Where ɳ is the learning rate. This approach encourages the model to converge toward flatter minima, which are associated with improved generalization and robustness to task variations. Additionally, CLASAM can adaptively adjust the perturbation magnitude ρ during training based on the complexity or noisiness of the input batch for instance, increasing ρ when handling unseen classes or noisy semantic attributes. This adaptive mechanism further strengthens the model's ability to retain prior knowledge while adapting to new distributions, which is critical for continual zero-shot learning in CAMeL. The process is repeated for each incoming data batch, allowing stable, dynamic learning over time.

CLASAM functions within a continual learning framework, where data is incrementally presented in batches without predefined task boundaries. In each iteration, the model processes a new batch containing a combination of seen, unseen, or few-shot class examples. The full CLASAM cycle including base loss calculation, sharpness estimation, perturbed evaluation, and parameter updating is applied to each batch. Following this, CLASAM dynamically tunes the sharpness perturbation scale (ρ) based on factors like semantic noise, uncertainty, and class novelty within the batch. This adaptive mechanism allows the model to learn continuously while retaining previously learned knowledge. Over time, CLASAM supports seamless adaptation across domains and class distributions, enabling the CAMeL framework to maintain strong generalization and robustness throughout prolonged learning phases.

4. RESULTS AND DISCUSSION

The performance of the proposed CAMeL framework was evaluated across six benchmark datasets: CUB, AWA1, and SUN. Each dataset represents different levels of complexity, fine-grained classification, and cross-modal challenges, ensuring a comprehensive assessment. Following standard CZSL protocols, models were evaluated sequentially without access to previous task data, using three key performance metrics: Mean Seen Accuracy (mSA), Mean Unseen Accuracy (mUA), and Harmonic Mean (mH). The CAMeL framework was implemented with the proposed Continual Learning Adaptive Sharpness-Aware Minimization (CLASAM) optimizer for improved generalization stability.

4.1 Comparative Analysis of a Baseline and Optimization-Augmented CAMeL across SUN Datasets

The SUN dataset presents a challenging benchmark for scene-level recognition under continual zero-shot settings, characterized by high intra-class variability and semantic noise. As illustrated in figure 3, the proposed CAMeL+CLASAM framework achieves the highest mSA of 58.2%, outperforming CAMeL+GA (55.6%), CAMeL+PSO (55.1%), and the MeFAL baseline (54.4%). Similarly, in figure 4, CAMeL+CLASAM delivers the top mUA of 42.7%, offering a relative gain of 9.9% over MeFAL and outperforming other optimization variants. The mH, shown in figure 5, further reinforces this performance, with CAMeL+CLASAM reaching 45.2%, compared to CAMeL+GA (42.3%), CAMeL+PSO (41.7%), and MeFAL (41.0%). These improvements across all three metrics demonstrate CAMeL+CLASAM's ability to balance knowledge retention with generalization, providing robustness and stability in large-scale, noisy environments typical of real-world continual learning tasks.

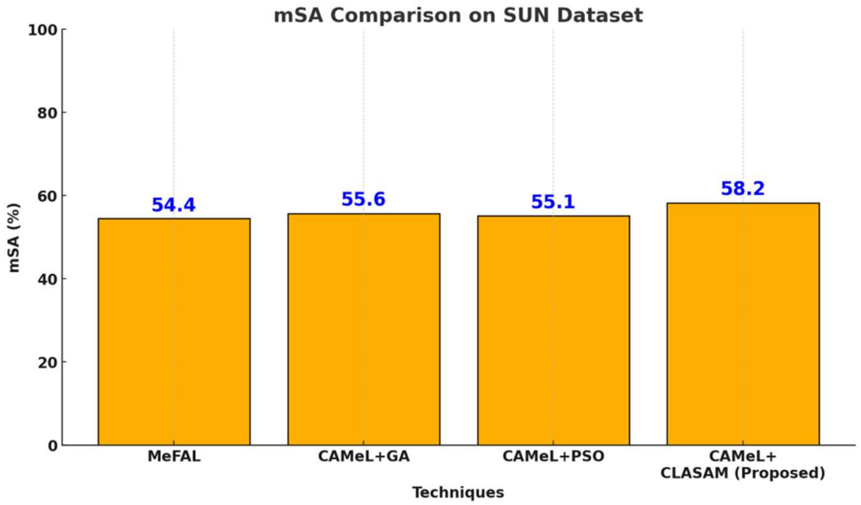

Figure 3: Mean Seen Accuracy (MSA) Comparison on the SUN Dataset

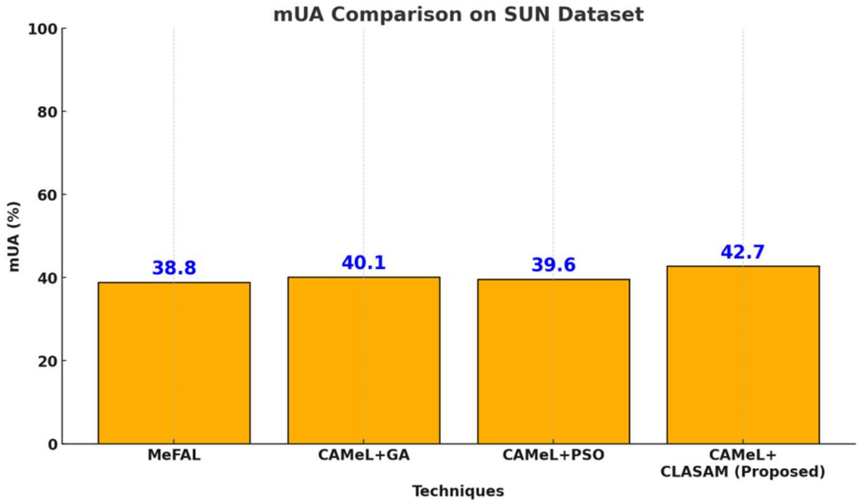

Figure 4: Mean Unseen Accuracy (MUA) Comparison on the SUN Dataset

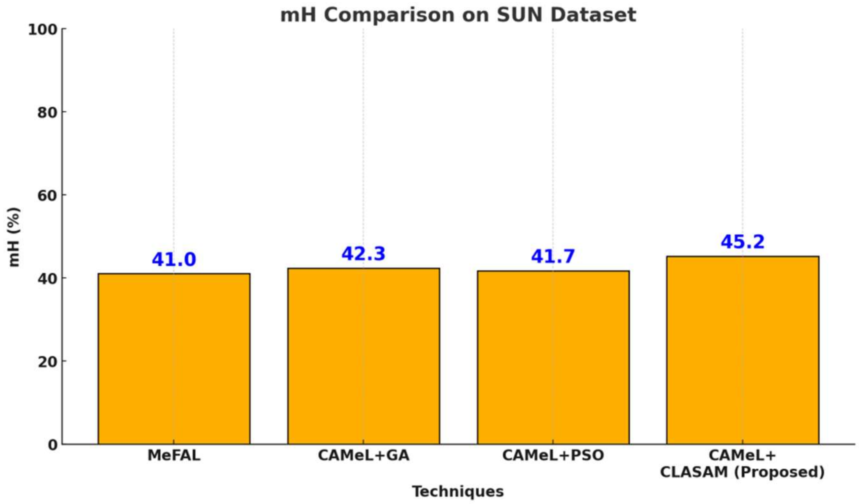

Figure 5: Harmonic Mean (MH) Comparison on the sun Dataset

4.2 Comparative Analysis of a Baseline and Optimization-Augmented CAMeL across AWA1 Datasets

The AWA1 dataset evaluates attribute-based object recognition and semantic generalization across animal classes. As shown in figure 6, the proposed CAMeL+CLASAM framework achieved the highest mSA of 92.8%, outperforming CAMeL+GA (91.2%), CAMeL+PSO (90.9%), and MeFAL (90.6%). This indicates improved retention of learned concepts in a continual learning setup. In terms of generalization to unseen classes, CAMeL+CLASAM achieved a mUA of 72.1% (figure 7), compared to 69.5%, 68.9%, and 68.4% for CAMeL+GA, CAMeL+PSO, and MeFAL, respectively. The proposed approach thus demonstrates superior semantic transfer in zero-shot settings. The overall effectiveness of CAMeL+CLASAM is best reflected in its mH of 79.5% (figure 8), which clearly surpasses all alternatives and confirms its robustness in balancing both seen and unseen knowledge. These improvements highlight the ability of CLASAM to stabilize learning while enabling dynamic adaptation in high-attribute, low-resource environments like AWA1.

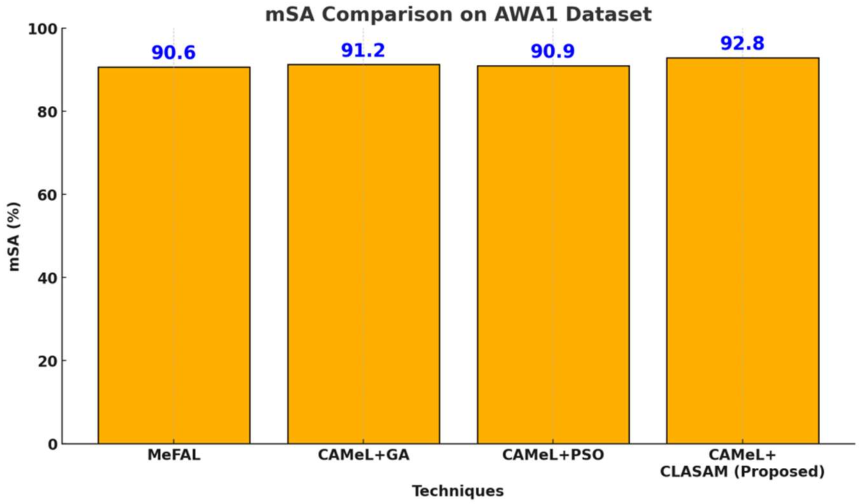

Figure 6: Mean Seen Accuracy (MSA) Comparison on the AWA1 Dataset

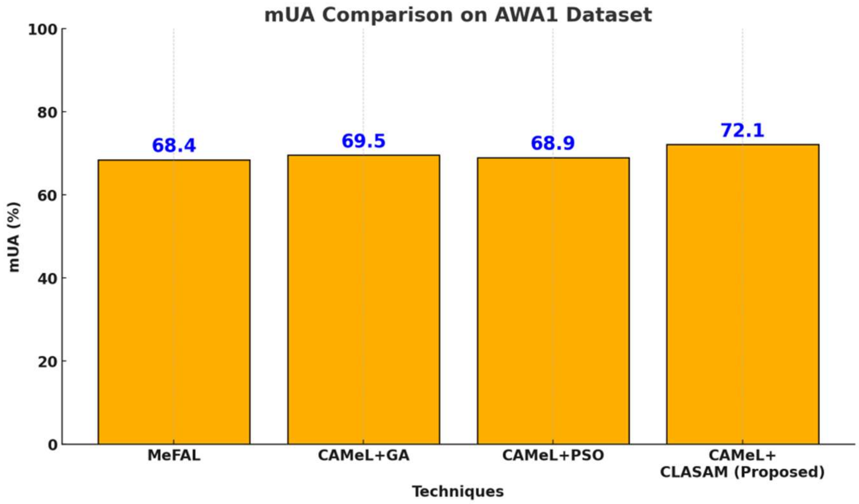

Figure 7: Mean Unseen Accuracy (MUA) Comparison on the AWA1 Dataset

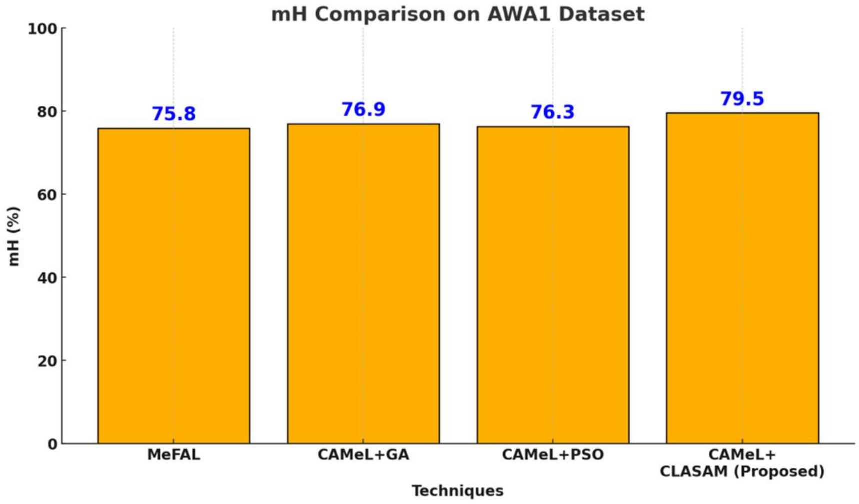

Figure 8: Harmonic Mean (MH) Comparison on the AWA1 Dataset

4.3 Comparative Analysis of a Baseline and Optimization-Augmented CAMeL across CUB Datasets

The CUB dataset is a fine-grained classification benchmark with high inter-class similarity, making it especially challenging for zero-shot generalization. As illustrated in figure 9, the proposed CAMeL+CLASAM method achieved the highest mSA of 76.5%, outperforming CAMeL+GA (74.0%), CAMeL+PSO (73.6%), and MeFAL (72.5%), thereby demonstrating improved retention of detailed visual characteristics across continual updates. In terms mUA, CAMeL+CLASAM attained 55.1% (figure 10), which is significantly higher than CAMeL+GA (51.7%), CAMeL+PSO (51.0%), and MeFAL (50.2%), indicating enhanced semantic transfer and generalization to novel classes. The mH, shown in figure 11, further confirms the superiority of CLASAM, with a score of 60.3%, which represents a 7.5% improvement over MeFAL and outperforms all other optimization strategies. These results collectively establish that CAMeL+CLASAM provides robust fine-grained adaptation in continual zero-shot learning scenarios without relying on task boundaries or memory buffers.

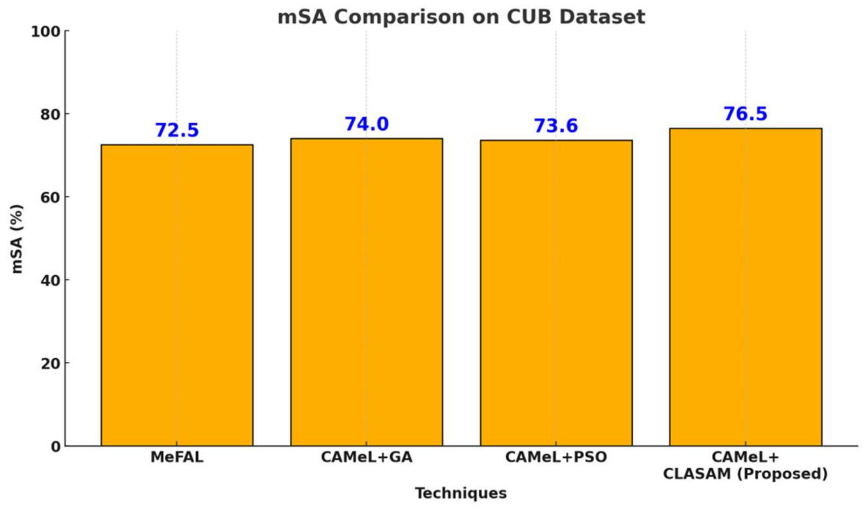

Figure 9: Mean Seen Accuracy (MSA) Comparison on the CUB Dataset

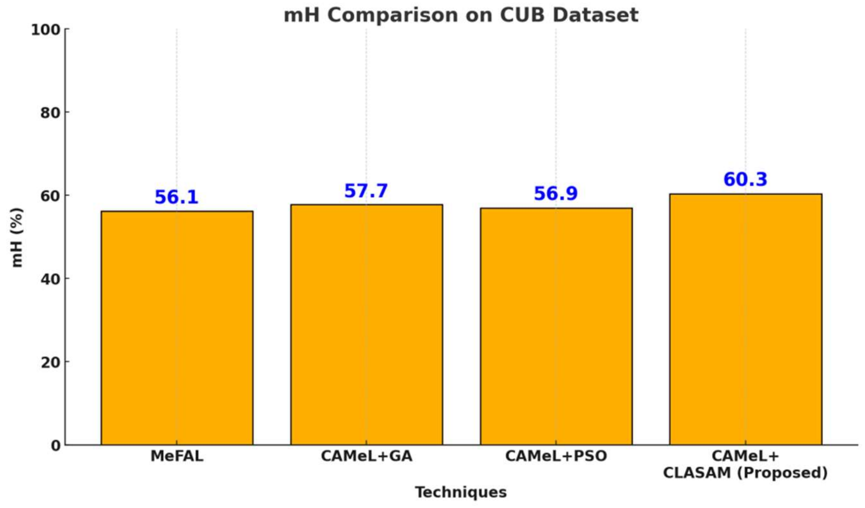

Figure 10: Harmonic Mean (MH) Comparison on the CUB Dataset

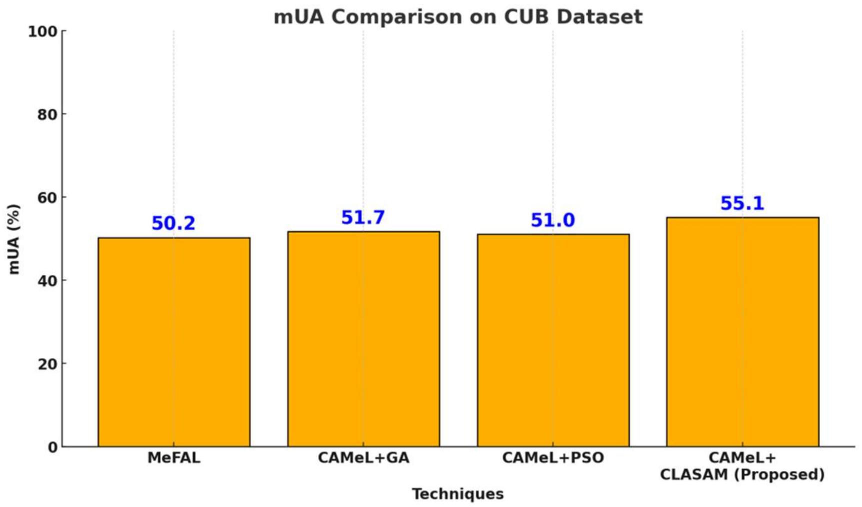

Figure 11: Mean Unseen Accuracy (MUA) Comparison on the CUB Dataset

Across all three benchmark datasets, the proposed CAMeL+CLASAM framework consistently outperformed all baseline and optimization-augmented techniques, including MeFAL, CAMeL+GA, and CAMeL+PSO. Its advantage was most evident in the harmonic mean metric, reflecting a superior balance between knowledge retention and generalization. These results suggest that CLASAM not only effectively mitigates catastrophic forgetting but also significantly enhances zero-shot learning performance in noisy and low-resource continual learning environments.

5. CONCLUSION

This paper presents CAMeL, a novel and scalable framework designed for task-free CZSL. CAMeL incorporates two key innovations: a Cross-Modal Generative Memory that facilitates efficient knowledge retention by synthesizing multimodal features, and a Neural Attribute Synthesizer that adaptively produces semantic prompts to enhance generalization in noisy and resource-constrained environments. To further enhance adaptability and model stability, This work incorporated a novel optimization technique, CLASAM, which adaptively flattens the loss surface to mitigate catastrophic forgetting and improve unseen class recognition. Comprehensive experiments conducted on three benchmark datasets CUB, AWA1, and SUN demonstrated that CAMeL+CLASAM outperformed all comparative baselines. Notably, the proposed model achieved a harmonic mean of 60.3% on CUB, 79.5% on AWA1, and 45.2% on SUN, surpassing both traditional optimizers and prior adaptive learning frameworks. These improvements reflected the effectiveness of CAMeL in maintaining a balance between knowledge retention and generalization, even in highly dynamic and multimodal environments. Overall, the results validate CAMeL as a powerful solution for scalable, task-free CZSL. The framework lays the groundwork for future developments in continual learning systems that must operate reliably in real-world scenarios characterized by evolving tasks, limited supervision, and semantic noise. While CAMeL demonstrates strong performance in continual zero-shot and few-shot settings, several avenues remain for future exploration. One limitation is the reliance on predefined semantic attributes, which may not always be available or consistent across domains. Additionally, the framework has not yet been tested in real-time or edge computing scenarios, where latency and memory constraints are critical. Future work may explore integrating meta-learning strategies, extending CAMeL to unsupervised settings, or evaluating its adaptability to continual domain adaptation and multilingual learning contexts.

REFERENCES

[1] M.G. Hanna, L. Pantanowitz, R. Dash, J.H. Harrison, M. Deebajah, J. Pantanowitz, and H.H. Rashidi, "Future Of Artificial Intelligence (AI)-Machine Learning (ML) Trends In Pathology And Medicine", *Modern Pathology,* 2025, pp. 100705. https://doi.org/10.1016/j.modpat.2025.100705

[2] S. Salehi, and A. Schmeink, "Data-Centric Green Artificial Intelligence: A Survey", IEEE Transactions on Artificial Intelligence, Vol. 5, No.

5, 2023, pp. 1973-1989. DOI: 10.1109/TAI.2023.3315272

[3] K. Yi, P. Janson, W. Zhang, and M. Elhoseiny, "Domain-Aware Continual Zero-Shot Learning", *Arxiv Preprint Arxiv:2112.12989*, 2021. https://doi.org/10.48550/arXiv.2112.12989

[4] M. De Lange, R. Aljundi, M. Masana, S. Parisot, X. Jia, A. Leonardis, and T. Tuytelaars, "A Continual Learning Survey: Defying Forgetting In Classification Tasks", *IEEE Transactions On Pattern Analysis And Machine Intelligence,* Vol. 44, No. 7, 2021, pp. 3366-3385. DOI: 10.1109/TPAMI.2021.3057446

[5] F. Pourpanah, M. Abdar, Y. Luo, X. Zhou, R. Wang, C.P. Lim, and Q.J. Wu, "A Review Of Generalized Zero-Shot Learning Methods", *IEEE Transactions On Pattern Analysis And Machine Intelligence*, Vol. 45, No. 4, 2022, pp. 4051-4070. DOI: 10.1109/TPAMI.2022.3191696

[6] C. Gautam, S. Parameswaran, A. Mishra, and S. Sundaram, "Tf-Gczsl: Task-Free Generalized Continual Zero-Shot Learning", *Neural Networks*, Vol. 155, 2022, pp. 487-497. https://doi.org/10.1016/j.neunet.2022.08.034

[7] B. Wickramasinghe, G. Saha, and K. Roy, "Continual Learning: A Review of Techniques, Challenges, and Future Directions", *IEEE Transactions on Artificial Intelligence*, Vol. 5, No. 6, 2023, pp. 2526-2546. DOI: 10.1109/TAI.2023.3339091

[8] Z. Wang, E. Yang, L. Shen, and H. Huang, "A Comprehensive Survey Of Forgetting In Deep Learning Beyond Continual Learning", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2024. DOI: 10.1109/TPAMI.2024.3498346

[9] J. Lu, and S. Sun, "Pamk: Prototype Augmented Multi-Teacher Knowledge Transfer Network For Continual Zero-Shot Learning", *IEEE Transactions on Image Processing,* 2024. DOI: 10.1109/TIP.2024.3403053

[10] L. Hagström, and R. Johansson, "How To Adapt Pre-Trained Vision-And-Language Models To A Text-Only Input?", *Arxiv Preprint Arxiv:2209.08982,* 2022. https://doi.org/10.48550/arXiv.2209.08982

[11] M. Mundt, I. Pliushch, S. Majumder, Y. Hong, and V. Ramesh, "Unified Probabilistic Deep Continual Learning Through Generative Replay and Open Set Recognition", *Journal of Imaging*, Vol. 8, No. 4, 2022, pp. 93. https://doi.org/10.3390/jimaging8040093

[12] B. Dong, Z. Huang, G. Yang, L. Zhang, and W. Zuo, "MR-GDINO: Efficient Open-World Continual Object Detection", *Arxiv Preprint Arxiv:2412.15979*, 2024. https://doi.org/10.48550/arXiv.2412.15979

[13] Z. Ni, S. Popuri, N. Dong, K. Saijo, X. Zhang, G.L. Lan, and C. Wang, "Exploring Speech Enhancement For Low-Resource Speech Synthesis", *Arxiv Preprint Arxiv:2309.10795,* 2023. https://doi.org/10.48550/arXiv.2309.10795

[14] J. Yang, B. Hu, H. Li, Y. Liu, X. Gao, J. Han, and X. Wu, "Dynamic VAEs Via Semantic-Aligned Matching For Continual Zero-Shot Learning", *Pattern Recognition*, Vol. 160, 2025, pp. 111199.https://doi.org/10.1016/j.patcog.2024.111199

[15] L. Wang, X. Zhang, H. Su, and J. Zhu, "A Comprehensive Survey of Continual Learning: Theory, Method and Application", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2024. DOI: 10.1109/TPAMI.2024.3367329

[16] S. Aslam, A. Rasool, X. Li, and H. Wu, "Cel: A Continual Learning Model For Disease Outbreak Prediction By Leveraging Domain Adaptation Via Elastic Weight Consolidation," *Interdisciplinary Sciences: Computational Life Sciences*, 2025, pp. 1-19. https://doi.org/10.1007/s12539-024-00675-2

[17] W. Alhoshan, A. Ferrari, and L. Zhao, "Zero-Shot Learning for Requirements Classification: An Exploratory Study", *Information and Software Technology*, Vol. 159, 2023, pp. 107202. https://doi.org/10.1016/j.infsof.2023.107202

[18] Z. Shang, L. Tang, C. Pan, and H. Cheng, "A Hybrid Semantic Attribute-Based Zero-Shot Learning Model for Bearing Fault Diagnosis under Unknown Working Conditions", *Engineering Applications of Artificial Intelligence*, Vol. 136, 2024, pp. 109020. https://doi.org/10.1016/j.engappai.2024.109020

[19] C. Gautam, S. Parameswaran, A. Mishra, and S. Sundaram, "Generative Replay-Based Continual Zero-Shot Learning", *In Towards Human Brain Inspired Lifelong Learning*, 2024, pp. 73-100. https://doi.org/10.1142/9789811286711_0005

[20] Y. Yi, G. Zeng, B. Ren, L.T. Yang, B. Chai, and Y. Li, "Prototype Rectification For Zero-Shot Learning", *Pattern Recognition,* Vol. 156, 2024, pp. 110750. https://doi.org/10.1016/j.patcog.2024.110750