Online Versatile Incremental Learning: Towards Class and Domain-Agnostic Adaptation at Any Time

Jae-Ho Lee[1], Min-Yeong Park[2*], Jun-Yeong Moon[1], Jung Uk Kim[2†], and Gyeong-Moon Park[1†]

[1] Korea University, Seoul, Republic of Korea
{jaeho-lee, moonjunyyy, gm-park}@korea.ac.kr
[2] Kyung Hee University, Yongin, Republic of Korea
{pmy0792, ju.kim}@khu.ac.kr

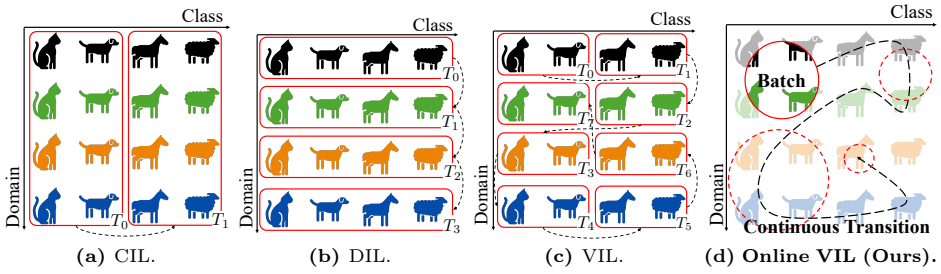
Abstract. Continual learning enables vision systems to adapt to ever-changing data distributions. Despite significant advances, existing approaches fail to capture continuous and concurrent shifts in classes and domains, a critical capability for real-world deployment. This work introduces **Online VIL (Online Versatile Incremental Learning)**, a novel scenario where class concepts and visual domains evolve simultaneously online without explicit boundaries. To better adapt to the challenges of such dynamic environments that more closely resemble real-world conditions, we propose a novel framework **TopFlow**, **Top**ology preservation with **Flow** matching representation that contains two complementary mechanisms: **Domain-agnostic Flow Matching (DFM)** and **Global Topology Preservation (GTP)**. DFM guides the model to have domain-agnostic representations by integrating the geodesic flow kernel into contrastive learning. In contrast, GTP maintains the global structure of the feature space without explicitly storing past examples. Our extensive experiments demonstrate that TopFlow effectively addresses the limitations of existing methods within the Online VIL scenario, achieving state-of-the-art performance in challenging Online VIL. The proposed methods suggest potential directions for building continual learning systems in realistic dynamic environments. Our implementation code is available at https://github.com/KU-VGI/Online-VIL.

Keywords: Online learning · Incremental learning · Real-world scenario

1 Introduction

Continual Learning (CL) [21,29,41,46,56–58,61] has gained increasing attention as deep learning moves closer to the real world, where data distributions evolve. A key challenge is catastrophic forgetting, where adapting to new knowledge disrupts

* Work done at Korea University.
† Corresponding Author.



(a) CIL. **(b)** DIL. **(c)** VIL. **(d) Online VIL (Ours).**

Fig. 1: Conceptual comparison of (a) Class Incremental Learning (CIL), (b) Domain Incremental Learning (DIL), (c) Versatile Incremental Learning (VIL), and (d) Online Versatile Incremental Learning (Online VIL, Ours). Red lines indicate the scopes observable at once, while black dashed lines depict transitions across time.

previous knowledge. To mitigate this, prior research has introduced distinct paradigms such as Class Incremental Learning (CIL) and Domain Incremental Learning (DIL), targeting a specific form of non-stationarity. More recently, Online Continual Learning (OCL) [18,54,59] has been proposed to handle streaming data under memory and single-pass constraints, including scenarios with blurry task boundaries [2,22,38]. However, most works still rely on CIL or DIL, limiting their ability to capture the full complexity of real-world dynamics.

A recently introduced Versatile Incremental Learning (VIL) [42] suggests a more realistic scenario where new tasks have a broader chance to evolve in both directions of the classes and domains, without prior knowledge. While VIL marks a step toward realism, it assumes discrete increments that provide implicit structural information about distribution shifts. In contrast, real-world environments exhibit continuous transitions without boundaries, offering no such organizational cues. Moreover, environmental change in reality is continuous and unpredictable, due to locally constrained information and multi-factor interactions. For instance, an autonomous driving system may suddenly face both a new object and a shift in weather or city conditions, requiring immediate (online) adaptation without knowing whether it concerns classes, domains, or both.

To this end, we introduce **Online V**ersatile **I**ncremental **L**earning (**Online VIL**), a new scenario that captures both the unpredictable heterogeneous shifts in an online manner. Online VIL is distinguished by reflecting the evolution of the natural information stream, characterized by unpredictable, gradual, and heterogeneous shifts that occur along multiple evolutionary trajectories. As shown in Figure 1, Online VIL allows flexible transitions across class and domain dimensions; each sequence presents distinct challenges without heuristic patterns. Consequently, Online VIL enables faithful evaluation of CL models and provides a foundation for systems that operate under real-world dynamics.

In the Online VIL scenario, adaptation to chaotic shifts in classes and domains without explicit access to prior inputs is crucial for distinguishing class-discriminative knowledge from domain-specific knowledge. Otherwise, models rely

on the spurious features of current distributions and tend to exhibit rapid forgetting and a lack of generalization. Through systematic layer-wise feature analysis of pre-trained Vision Transformers, we observe that early layers predominantly capture domain-specific patterns (e.g., texture, lighting), while deeper layers encode class-discriminative features and more abstract semantic representations. This motivates a novel **Domain-agnostic Flow Matching (DFM)** technique, which aligns features with the intrinsic geometry of pre-trained knowledge through reconceptualized geodesic flow kernel [12] while mitigating domain-specific shifts.

In addition to class and domain-agnostic alignment, preserving the global topology of learned features is essential for maintaining semantic continuity across evolving tasks. Conventional objectives often fail to maintain the structural relationships of feature space because they shrink the occupation of missing classes in the feature space. To address this, we propose **Global Topology Preservation (GTP)**, which preserves invariant geometric configurations in feature space, enabling robust knowledge retention without requiring complete class coverage.

Integrating these components, we introduce **Top**ology preservation with **Flow** matching representation (**TopFlow**), a novel framework designed for Online VIL. With recognition of DFM and GTP regularization for geometry, TopFlow ensures robust adaptation to unpredictable shifts. To evaluate its effectiveness, we conduct extensive experiments in Online VIL and observe that TopFlow consistently outperforms existing state-of-the-art methods across multiple benchmarks. Our contributions are summarized as follows:

- We introduce **Online VIL**, a realistic scenario for evaluating continual learning where class and domain distributions evolve continuously and jointly with ambiguous task boundaries.
- We reveal a novel role of the pre-trained ViT layer that encodes class and domain knowledge. Leveraging this insight, we propose Domain-agnostic Flow Matching (**DFM**) to learn domain-agnostic representations by integrating the geodesic flow kernel into contrastive learning.
- We propose Global Topology Preservation (**GTP**), a mechanism for maintaining knowledge representations using feature topologies without explicit memory of previous inputs.
- We demonstrate that the proposed **TopFlow** significantly outperforms existing state-of-the-art methods through comprehensive experiments in our challenging Online VIL scenario.

2 Related Work

Online Continual Learning. Online Continual Learning (OCL) [5, 15] has emerged as a pragmatic paradigm that reflects the challenges of real-world settings, where data arrive as continuous streams and models must operate under minimal batch sizes, single-pass constraints, and strict computational and memory limitations. Traditionally, OCL methods rely on replay buffers [37, 44] to store a small subset of past data, thereby mitigating catastrophic forgetting while learning new tasks. Recent advances explore leveraging prototypes [59], replay-free

strategies [60], and pre-trained models with prompt [38]. While effective in class or domain increments, this dependence on memory limits scalability and realism. Methods addressing more realistic scenarios with blurry or ambiguous task boundaries have emerged [2,22]. However, the replay methods require growing memory in proportion to task diversity. Also, blurry boundary methods still assume either class-only or domain-only shifts, failing to capture the heterogeneous evolution of a realistic data stream. Therefore, we propose Online Versatile Incremental Learning (Online VIL). This scenario exposes models to unpredictable shifts in both classes and domains while enforcing online constraints that limit memory and multi-pass access.

Geodesic Flow Kernel. The Geodesic Flow Kernel (GFK) [12,14] has been widely used in unsupervised domain adaptation to align feature distributions between a predefined source and target domain. Conventional applications approximate the geodesic on the Grassmannian manifold between two static domains, which requires that the data of the source and target domains are fully available and static. In contrast, Online VIL presents unique challenges: domain shifts occur continuously and unpredictably, and data arrive sequentially, making it impossible to estimate the flow offline. To address these challenges, we reformulate the approach from GFK and design a novel Domain-Agnostic Flow Matching (DFM). Unlike traditional geometry estimation in feature space, DFM is designed for sequentially arriving data and evolving domains without relying on holistic data access. Through both empirical and theoretical analysis of feature geometry, DFM enables online alignment on feature geometry in dynamic conditions, avoiding computationally expensive higher-order manifold computations. This design fundamentally extends the applicability and purpose of GFK, enabling robust domain-agnostic feature matching in the OCL.

Feature Topology. Several works have leveraged the topology of feature space to mitigate catastrophic forgetting in sequential learning. [49] employs elastic Hebbian graphs to preserve neighborhood relationships during incremental updates, while [50] uses self-organizing maps to identify representative feature points and restrict their displacement. More recent approaches, such as [33] and [55], maintain pair-wise instance similarity or local topological relations by decomposing the global structure, further reducing forgetting across tasks. However, these methods are limited to offline CL, requiring full access to past samples to construct the feature topology. Furthermore, existing topology preservation methods rely on complete semantic information across all classes. In online scenarios where only partial class information is available at each time step, these methods suffer from incomplete topology construction, resulting in suboptimal feature space organization. In contrast, our work proposes a Global Topology Preservation (GTP) strategy that maintains structural knowledge without requiring the storage of previous data.

3 Method

3.1 Problem Setup: Online Versatile Incremental Learning

In real-world scenarios, data distributions gradually evolve, often exhibiting significant variability across multiple dimensions such as spatial domains, semantic classes, and temporal shifts. We propose a novel scenario termed Online Versatile Incremental Learning (Online VIL), which simulates these properties by constructing a continuous data stream with stochastically varying distributions. Given a dataset with $n_{\mathcal{D}}$domains and $n_{\mathcal{C}}$ classes, we define the product category space as follows:

$$\mathcal{X} = \bigcup_{i=1}^{n_{\mathcal{D}}} \mathcal{D}_i, = \bigcup_{j=1}^{n_{\mathcal{C}}} \mathcal{C}_j, = \bigcup_{i=1}^{n_{\mathcal{D}}} \bigcup_{j=1}^{n_{\mathcal{C}}} \mathcal{K}_{i,j}, \quad \text{where} \quad \mathcal{K}_{i,j} = \mathcal{D}_i \cap \mathcal{C}_j, \tag{1}$$

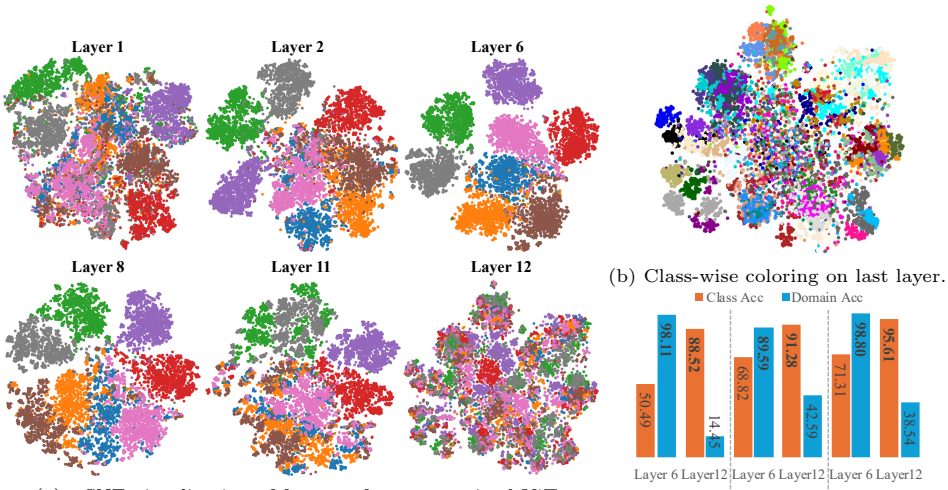
where $\mathcal{D}_i$ is a set of samples in a domain i and $\mathcal{C}_j$ is in a class j, and $\mathcal{K}_{i,j}$ is samples that belong to domain i and class j. Machine learning typically assumes the data $\mathcal{X}$ as a lower-dimensional manifold embedded in a high-dimensional space [3]. In this context, we can conceptualize the data stream as a trajectory through the joint domain-class space, where each timestep t provides an observation window $\mathcal{V}_t \subset \mathcal{X}$ which is a disjoint open neighborhood of data $\mathcal{X}$. This geometric interpretation naturally leads to our manifold-based approach in the subsequent technical development.

Inspired by Si-Blurry [38], we divide categories into disjoint sets ($\mathcal{K}_{\text{disjoint}}$) with clear distributional boundaries and blurry sets ($\mathcal{K}_{\text{blurry}}$) with deliberately abstracted distributions.

Algorithm A.1 in Appendix details our task construction process, which proceeds in six main steps: (*i*) *category partitioning* to establish variable distributional clarity, (*ii*) *sample extraction* to create diverse distributional patterns, (*iii*) *sample redistribution* to simulate realistic category overlap, (*iv*) *randomized task assignment* to eliminate artificial task boundaries, (*v*) *task construction* with varying characteristics, and (*vi*) *batch generation* with constrained visibility windows. The Online VIL scenario distinguishes itself from traditional CL scenarios through three key characteristics:

1. **Locally Limited Visibility.** At any time step, the model observes only a small fraction in both number of samples ($|\mathcal{V}_t| \ll |\mathcal{D}|$) and categories ($\exists \mathcal{K}_{i,j} : x \in \mathcal{K}_{i,j} \wedge x \notin \mathcal{V}_t$), reflecting real-world constraints on data accessibility, combines both spatial and temporal restrictions.
2. **Continuous Smooth Variation.** The variation is smooth and continuous over time, making it hard to distinguish the distribution shift without a global context.
3. **Dynamic Distribution.** Each category appears and disappears gradually. The variance across spatial (domain appearances), semantic (class properties), and temporal (distribution evolution) dimensions requires simultaneous adaptation to new patterns and retention of previously acquired knowledge.

This formulation demands adaptation mechanisms that extract meaningful patterns despite high variance and constrained observability. Traditional CL methods,



(a) t-SNE visualization of features from pre-trained ViT.

(b) Class-wise coloring on last layer.

(c) Linear Probing results.

Fig. 2: Analysis of layer-wise knowledge on pre-trained ViT for data with changing distributions (CORe50). (a) t-SNE visualization of features from certain layers of the pre-trained ViT. The color indicates the domain. (b) The t-SNE visualization of the last layer with class-wise color. (c) Accuracy of linear probing for class and domain classification from different layers.

which assume either domain stability or class stability, fail to resemble these fundamental challenges in real-world perception systems.

3.2 Domain-agnostic Flow Matching

The Online VIL scenario, with limited visibility, multi-dimensional variance, and dynamic distribution, poses challenges beyond traditional CL. While the original VIL work measured feature similarity, it did not explicitly examine *how* domain and class signals are encoded across the backbone under noisy and concurrent class/domain shifts. This matters because the one-pass constraint and the lack of rehearsal make representations highly sensitive to the local batch composition, and thus prone to overfitting transient domain cues. To better understand this, we analyze the representations of *frozen* pre-trained Vision Transformers, which have become the standard backbone in recent CL research [11, 46, 56–58]. Although not trained under Online VIL, a frozen backbone provides a stable proxy, reflecting common practice in CL of freezing the ViT and updating only a small number of parameters (e.g., prompts). Its hierarchical representations largely dictate how domain-specific variations and class-level semantics interact. Motivated by this, we conduct a layer-wise analysis, expecting deeper layers to align more closely with class semantics, as their outputs directly drive the classification head.
We perform a t-SNE study and linear probing to investigate this. Our t-SNE visualizations (Figures 2a, 2b) show that intermediate features group by *domain*,

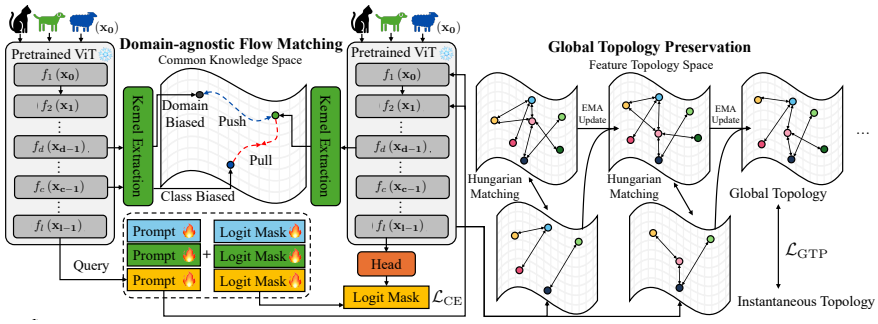


Fig. 3: Architecture overview. TopFlow comprises two components: Domain-agnostic Flow Matching (DFM) and Global Topology Preservation (GTP). DFM promotes domain-agnostic representation accumulation, while GTP preserves the global feature topology.

while final features group by *class*. Linear probing (Figure 2c) further confirms stronger domain discrimination in intermediate layers and stronger class discrimination in the final layer. These findings reveal a *structural representation gap* in Online VIL, motivating our proposed **Domain-agnostic Flow Matching (DFM)**: a geometry-informed contrastive loss that pulls the *adaptable intermediate* representation toward the *semantic* geometry of the final layer while pushing it away from domain-biased directions.

From Geodesic-Flow Intuition to Layer-Wise Comparison Geometry. The Geodesic Flow Kernel (GFK) [12] suggests that features from a distribution can be represented by a subspace on the Grassmann manifold, and similarity can be computed through a comparison geometry defined over such subspaces. In Online VIL, however, explicitly estimating *inter-domain* subspaces is often ill-posed due to continuous, mixed, and unlabeled domain evolution. Instead, we leverage the empirical layer-wise gap in Fig. 2 and treat *layers* as the unit of geometry: intermediate features are typically more domain-sensitive, while final features are more class-semantic. Accordingly, we extract a *batch-wise common comparison geometry* between intermediate and final representations, and use it to normalize similarities in a contrastive objective. *Residual-layer dynamics and local approximation.* For convenience, we consider architectures with matched feature dimensionality across layers, such as ViTs [8]. With a cascade of functions with residual connections:

$$\boldsymbol{h}_0 = \boldsymbol{x}, \quad \boldsymbol{h}_n = f_n(\boldsymbol{h}_{n-1}) + \boldsymbol{h}_{n-1}, \quad n = 1, 2, \ldots, l, \tag{2}$$

where $f_i : \mathbb{R}^{b\times d} \to \mathbb{R}^{b\times d}$ is a function of the i-th layer, and $\boldsymbol{h}_l \in \mathbb{R}^{b\times d}$ is the last layer feature with batch size b and feature dimension d. With the local neighborhood $\mathcal{V}_t$ on a Riemannian manifold as mentioned in Section 3.1, we can take a first-order approximation with the infinitesimal variation of feature $\boldsymbol{h}_l$, as detailed in Equation A.3 and A.5. Then, the inner product between h_n and h_l

can be written as

$$\langle \boldsymbol{h}_n, \boldsymbol{h}_l \rangle = \int_X \boldsymbol{h}_n^T \boldsymbol{h}_l d\boldsymbol{x} = \mathbb{E}_X \left[(\bar{\boldsymbol{h}}_n + \delta \boldsymbol{h}_n)^T (\bar{\boldsymbol{h}}_l + \delta \boldsymbol{h}_l) \right], \tag{3}$$

where $\bar{\boldsymbol{h}}_n = \mathbb{E}_{\boldsymbol{x} \in \mathcal{V}_t}[\boldsymbol{h}_n]$ denotes the local mean feature over the neighborhood $\mathcal{V}_t$, and $\delta \boldsymbol{h}_n = \boldsymbol{h}_n - \bar{\boldsymbol{h}}_n$ is the corresponding local deviation.

Common Subspace Extraction and the Role of U. To enable a geometry-aware comparison, we extract a *common subspace* from the combined space of $\boldsymbol{h}_n$ and $\boldsymbol{h}_l$ by stacking $H = \left[\boldsymbol{h}_n^T \; \boldsymbol{h}_l^T\right]^T = U \Sigma V^T$. As detailed in Equation A.3, projecting onto the common subspace U provides an orthogonal basis that summarizes correlated directions shared across layers within the local neighborhood. Based on observation in Figure 2, we interpret the projected components as:

- $\delta \boldsymbol{h}_n^T \delta \boldsymbol{h}_n$ represents the residual component ($\boldsymbol{u}_{\text{residual}}$), dominated by layer-local (often domain-sensitive) variation;
- $\delta \boldsymbol{h}_n^T \delta \boldsymbol{h}_l$ and $\delta \boldsymbol{h}_l^T \delta \boldsymbol{h}_n$ represent push-forward induced by transportation ($\boldsymbol{u}_{\text{push}}$), capturing cross-layer transport toward semantic features;
- $\delta \boldsymbol{h}_n^\top \left(\Phi_n^l\right)^\top \Phi_n^l \delta \boldsymbol{h}_n$ provides metric of the push-forward transportation ($\boldsymbol{u}_{\text{metric}}$), reflecting local metric/curvature information.

While component-wise basis isolation is non-trivial in an online mixed stream, the shared subspace U provides a batch-wise geometry that emphasizes correlated cross-layer directions and suppresses batch-specific noise. This geometry is used for similarity normalization in DFM.

Domain-agnostic Flow Matching Loss. With this insight, we guide intermediate representations away from domain-specific geometry and toward class-discriminative geometry. Let $\boldsymbol{h}_n^*$ be a perturbed function of $\boldsymbol{h}_n$ with continuous deformation (e.g., induced by prompts); locally, we apply the same U extracted above. For a sample $\boldsymbol{x}_{(i)}$ in batch $\boldsymbol{X}$, let the intermediate feature $\boldsymbol{h}_{n,(i)} \in \boldsymbol{H}_n$, last-layer feature $\boldsymbol{h}_{l,(i)} \in \boldsymbol{H}_l$ from the frozen backbone model, and intermediate features from the perturbed function $\boldsymbol{h}_{n,(i)}^* \in \boldsymbol{H}_n^*$.

Positive/Negative Design: Semantic Pull and Domain-Biased Push. We use the final-layer feature as a semantic anchor and set $\boldsymbol{H}_{(i)}^+ = \left\{\boldsymbol{h}_{l,(i)}\right\}$ as the positive set. To push $\boldsymbol{h}_{n,(i)}^*$ away from domain-sensitive intermediate regions (Fig. 2a), we use frozen intermediate features as *domain-biased negatives*. Other samples' final-layer features are also included to prevent collapse and preserve instance-level discrimination. The negative set is

$$\boldsymbol{H}_{(i)}^- = \left\{\boldsymbol{h}_{n,(j)} \mid \boldsymbol{h}_{n,(j)} \in \boldsymbol{H}_n\right\} \cup \left\{\boldsymbol{h}_{l,(j)} \mid i \neq j \wedge \boldsymbol{h}_{l,(j)} \in \boldsymbol{H}_l\right\}. \tag{4}$$

We propose a contrastive loss formulated as follows:

$$\mathcal{L}_{\text{DFM}} = -\frac{1}{M} \sum_{m=1}^{M} \log \frac{\sum_{\boldsymbol{h}_{n,(m)}^+ \in \boldsymbol{H}_{(m)}^+} \exp\left(\cos(\boldsymbol{h}_{(m)}^* U, \boldsymbol{h}_{(m)}^+ U)/\tau\right)}{\sum_{\boldsymbol{h}_{n,(m)}^+ \in \boldsymbol{H}_{(m)}^+} \exp\left(\cos(\boldsymbol{h}_{(m)}^* U, \boldsymbol{h}_{(m)}^- U/\tau\right)}, \tag{5}$$

where cos is cosine similarity and τ is a temperature, while dividing by norm reduces the effect of the amplitude of the feature. The further details of the derivation and procedure are provided in Appendix A.2 and Algorithm A.2.

3.3 Global Topology Preservation

One significant challenge in Online VIL is that the model must train from batches that only partially represent the overall distribution. This issue is particularly acute when input batches exhibit non-stationary class-domain compositions. This challenge is particularly acute in the Online VIL scenario, where input batches exhibit non-stationary class-domain compositions. To address this representational instability and to ensure topological coherence of the feature space, we introduce **Global Topology Preservation (GTP)**, a novel approach designed to maintain semantic structural integrity across temporally evolving data streams. The effective information content of batch B_t at layer $\boldsymbol{h}_l(\boldsymbol{x})$ can be measured by the rank of its empirical covariance matrix:

$$C_t = \mathbb{E}_{\boldsymbol{x}\in B_t}\left[(\boldsymbol{h}_l(\boldsymbol{x}) - \mu_t)(\boldsymbol{h}_l(\boldsymbol{x}) - \mu_t)^T\right], \quad \mu_t = \mathbb{E}_{\boldsymbol{x}\in B_t}\left[\boldsymbol{h}_l(\boldsymbol{x})\right], \tag{6}$$

when certain classes are absent, C_t captures only partial information about the global feature distribution, leading to rank deficiency. When batch B_t contains samples from only $k < C$ classes, the rank of the gradient is limited to k; the gradient provides not only the construction of decision boundaries for the present classes but also distorts the feature space by contracting around them. This rank deficiency yields incomplete feature representations and biases gradients toward observed classes, causing the feature space to contract around present classes while neglecting absent ones. This issue persists even with a prototype-based classifier [47], which assumes each class is independent, thereby removing the absent rank into a zero-eigenvalue space.

To circumvent these limitations while preserving global topological properties, GTP constructs a surrogate representation of the feature manifold through two key components: a set of k global prototypes $\{\bar{p}_g^j\}_{j=1}^k$, and a set of global relationship vectors $\{\bar{r}_g^{j,l}\}_{j,l=1}^k$. For each incoming batch B_t, we derive batch-specific prototypes $\{p_b^i\}_{i=1}^k$ from the final layer features using FINCH clustering [45]. We employ an exponential moving average (EMA) update strategy to ensure smooth temporal evolution of global prototypes while mitigating batch-to-batch fluctuations. Once the correspondence is established, each global prototype $\bar{p}_g$ is updated with the batch prototype p_b:

$$\bar{p}_g^{\pi^*(i),\text{new}} = (1-\alpha)\bar{p}_g^{\pi^*(i),\text{old}} + \alpha p_b^i \quad \text{for } i = 1,\ldots,k, \tag{7}$$

where α controls EMA update rate, balancing stability and adaptability. We find optimal assignment π^* by solving: $\pi^* = \arg\min_{\pi\in S_k}\sum_{i=1}^k \|p_b^i - \bar{p}_g^{\pi(i)}\|$ using Hungarian algorithm [24].

The relationships between its constituent elements fundamentally characterize the topological structure. To capture these structural properties, We introduce a learnable mapping function $\phi : \mathbb{R}^{2d} \to \mathbb{R}^m$, where concatenation preserves both individual prototype and relative positioning. We define batch-specific relationship vectors $r_b^{i,j} = \phi([p_b^i; p_b^j])$ and global relationship vectors $\bar{r}_g^{i',j'} = \phi([\bar{p}_g^{i'}; \bar{p}_g^{j'}])$. The GTP loss then aligns these relationship structures as follows:

$$\mathcal{L}_{\text{GTP}} = \sum_{i=1}^{k}\sum_{\substack{j=1\\ j\neq i}}^{k} D(r_b^{i,j}, \bar{r}_g^{\pi^*(i),\pi^*(j)}), \tag{8}$$

where D represents the cosine distance metric, this formulation encourages consistent pairwise relationships between semantic prototypes, effectively preserving the global topological structure without explicitly computing spectral properties.

Since the number of global prototypes $\bar{p}_g$ is smaller than the number of samples in each batch, these prototypes deliberately capture a coarse-grained representation, reflecting the intended vagueness under limited observations. Each prototype is updated considering its relationships with all other prototypes. These prototypes cover the feature space of several semantic classes, which reduces the distortion of the feature space from the concept without samples. Additionally, the EMA update strategy stabilizes the prototypes over time and removes outdated information from previous batches. As shown in Figure A.3 and A.4 in Appendix, this effectively filters out the unseen class information in recent batches, performing a role similar to a short-term memory. This mechanism provides a computationally efficient solution to maintaining representational stability in non-stationary environments, complementing the domain-invariance properties induced by DFM.

4 Experiments

This section evaluates and compares proposed TopFlow against state-of-the-art methods. Section 4.1 describes the experimental setup, including datasets, baselines, and evaluation metrics. Section 4.2 presents extensive quantitative results, demonstrating the effectiveness of TopFlow across multiple benchmarks. To further analyze TopFlow, Section 4.3 provides in-depth analyses, including ablation studies, to isolate and assess the contribution of each component in TopFlow. Further details about the experiments are provided in the Appendix B.

4.1 Experimental Setup

Datasets. We conducted experiments on three benchmarks, including iDigits [53], CORe50 [34], and CLEAR100 [31], for which it is possible to construct Online VIL scenarios that can cause a significant shift in distribution by clearly distinguishing both classes and domains. We split the CORe50 and CLEAR100 datasets into 10 tasks for the task construction, and 5 for the iDigits dataset.

Baselines. We compared our proposed method with traditional naive baselines and the latest state-of-the-art methods. First, we set the lower bound as the usual supervised sequential fine-tuning result (FT) and the upper bound as the usual supervised joint fine-tuning result. Then, we compared our proposed method with replay-based methods such as ER [44], Rainbow Memory (RM) [2], CLIB [22], and CBA [54], regularization-based method EWC [21], LwF [30], SLCA [61], DYSON [18], OCM [16], OnPro [59], PEC [60], S6MOD [32], DUCT [63] and prompt-based CODA-P [46], ICON [42] and MVP [38].

Implementation Details. We used the MVP [38] as our baseline for model and experimental setups. We used Adam optimizer with a learning rate of 5e-3, and implemented with a batch size of 64. As a mapping function ψ which maps

Table 1: Experimental results with proposed Online VIL scenarios. We used bold and underlined as brief indications of the best and the second best, respectively.

Method	iDigits		CORe50		CLEAR100	
	A_{AUC}	A_{Last}	A_{AUC}	A_{Last}	A_{AUC}	A_{Last}
Upper-bound	-	87.18±0.13	-	91.66±0.25	-	94.36±0.28
Lower-bound	13.45±0.62	12.71±3.34	3.39±0.10	3.24±1.09	2.38±0.32	2.66±0.55
EWC [21]	20.07±2.84	14.67±3.61	18.60±4.64	16.07±1.56	23.61±3.11	19.93±1.33
LwF [30]	19.61±3.50	15.38±1.04	25.27±3.77	21.97±4.22	23.80±2.24	21.70±4.19
CODA-P [46]	23.96±4.74	20.62±3.47	54.06±5.32	48.88±2.92	28.82±5.77	25.61±3.52
SLCA [61]	35.81±3.98	24.88±2.82	33.49±4.47	27.36±2.06	32.16±2.28	31.70±1.13
PEC [60]	34.77±3.02	28.01±2.69	51.35±4.39	46.95±2.28	53.93±3.20	51.66±2.09
ICON [42]	33.60±2.16	30.63±2.77	49.42±3.29	45.15±2.94	59.38±2.46	58.60±2.57
S6MOD [32]	31.18±1.62	30.42±2.55	52.33±2.74	47.20±3.15	59.60±2.79	57.47±2.48
DUCT [63]	36.46±3.48	30.94±2.41	53.66±4.37	47.24±1.28	66.79±2.94	65.46±3.40
MVP [38]	38.29±5.74	31.05±3.15	58.30±4.48	52.84±1.17	79.73±3.59	77.11±2.33
TopFlow (Ours)	**48.52±1.25**	**32.18±1.01**	**64.51±2.50**	**66.20±4.18**	**87.12±0.01**	**80.64±2.67**

prototypes into a relation vector, we used a simple 2-layer MLP function. The hidden dimension of ψ is 64 in CORe50, and 32 in CLEAR100. The size of the dimension of the relation vector m is 10. For the EMA update of global feature topology, a decay factor of 0.99 was used. Experiments were conducted under assumption of the memory-free setting, we adopted naïve reservoir memory for the experiments with memory buffer. We conducted experiments with 3 random seeds, and note that using more seeds (e.g., 10 runs) also yields consistent results. Details are described in the Appendix Section C.5 and Table A.9.

Evaluation Metrics. To evaluate online learning performance, we employed two metrics: A_{AUC} and A_{Last} [22]. The A_{AUC} metric quantifies performance under anytime inference, where inference queries may occur at arbitrary points during training as new classes are encountered. Conversely, A_{Last} assesses inference accuracy after training. In real-world applications, models must deliver reliable predictions on demand, regardless of training stage. Thus, A_{AUC} and A_{Last} provide a robust framework for benchmarking online learning performance.

4.2 Experimental Results

We conducted extensive experiments in the proposed Online VIL scenario, and the results are summarized in Table 1 and Table 2. As shown in Table 1, under the non-replay setting, TopFlow consistently outperforms existing methods in both A_{AUC} and A_{Last}. These results demonstrate that our approach mitigates catastrophic forgetting while enabling continual adaptation to the input stream. Moreover, in Table 2, the introduction of replay memory generally improves performance, but the proposed TopFlow consistently outperforms all baselines. Surprisingly, in replay-buffer settings, we observe that naïve methods such as Experience Replay (ER) and earlier methods tend to achieve the best performance, except for our proposed method. This suggests that most existing

Table 2: Results of OnlineVIL scenarios using replay buffer sizes 500 and 2000.

Buffer Size	Method	iDigits		CORe50		CLEAR100	
		A_{AUC}	A_{Last}	A_{AUC}	A_{Last}	A_{AUC}	A_{Last}
500	ER [44]	59.43±6.24	48.70±2.51	74.77±4.85	72.25±2.27	73.92±3.93	71.49±3.20
	RM [2]	55.02±5.37	51.73±2.05	81.06±3.90	70.41±3.17	72.42±4.64	72.93±2.05
	CLIB [22]	57.38±4.16	52.63±3.38	75.06±5.81	71.93±1.06	68.39±5.25	66.92±1.52
	OCM [16]	57.40±3.60	52.88±2.52	75.29±3.10	72.66±1.93	77.80±3.25	75.10±2.42
	CBA [54]	58.05±4.39	54.28±3.07	81.92±4.04	81.02±1.39	75.26±4.08	74.47±1.82
	OnPro [59]	46.92±4.82	44.86±1.63	71.39±4.03	70.92±2.24	81.36±4.99	77.46±1.53
	DYSON [18]	42.31±3.11	38.18±3.56	62.92±5.61	60.72±2.16	66.56±4.65	65.62±2.74
	MVP-R [38]	48.29±3.73	40.97±2.15	83.26±5.16	80.16±1.03	87.82±3.17	85.65±2.18
	TopFlow (Ours)	**62.48±4.17**	**56.48±1.52**	**85.16±0.84**	**91.14±0.05**	**91.67±0.03**	**90.12±1.00**
2000	ER [44]	61.72±5.12	57.81±1.47	79.44±5.17	76.61±3.91	83.51±4.61	81.03±3.59
	RM [2]	43.96±6.18	41.23±3.62	82.42±3.03	79.84±2.19	84.73±4.63	81.49±2.26
	CLIB [22]	51.96±3.81	46.06±2.34	84.58±4.26	81.63±2.51	85.62±5.05	83.25±1.62
	OCM [16]	57.52±5.08	50.60±3.34	84.92±4.03	83.24±2.72	84.26±4.82	82.91±3.56
	CBA [54]	60.02±4.52	55.04±2.82	85.16±3.28	83.49±3.20	88.02±3.80	85.94±2.36
	OnPro [59]	54.82±4.95	51.03±3.77	81.35±5.51	78.09±3.91	88.83±5.16	86.11±3.57
	DYSON [18]	46.74±4.41	43.38±4.28	51.21±3.71	49.29±1.74	57.05±4.18	55.48±3.43
	MVP-R [38]	52.14±2.99	47.74±2.36	87.33±3.37	82.39±1.10	89.48±1.71	88.93±1.18
	TopFlow (Ours)	**65.45±3.84**	**60.39±4.28**	**87.56±0.82**	**92.24±0.15**	**93.55±0.02**	**92.97±0.65**

Table 3: Ablation study for DFM and GTP on CORe50.

DFM	GTP	A_{AUC}	A_{Last}
Baseline		58.30	52.84
✓		64.13	64.57
	✓	63.95	65.28
✓	✓	**64.51**	**66.20**

Table 4: The Effectiveness of TopFlow in Standard CL Scenarios.

Method	Accuracy CIL, CODA-P	DIL, S-Prompt
Baseline	84.17	82.96
+ DFM	84.19	85.36
+ GTP	85.52	86.88
+ DFM, GTP	**86.04**	**87.50**

Table 5: Ablation of layer selection in the DFM loss (w/o GTP).

Layer	A_{Last}
Baseline	52.84
(0, 5)	49.92
(5, 10)	52.98
(5, 10), (6, 11)	53.49
(6, 11) (Ours)	**54.82**

online-incremental learning methods struggle in realistic Online VIL scenarios where distribution shifts are frequent and task boundaries are unclear. In contrast, our proposed TopFlow maintains strong performance across all settings, confirming its robustness. Furthermore, TopFlow achieves a more stable accuracy trajectory over time, with fewer drastic performance drops between tasks. This indicates that DFM and GTP help smooth the learning process by leveraging more structured representations and effective feature alignment.

4.3 Ablation Studies and Analysis

Ablation Study. Table 3 demonstrates the effectiveness of each component in our TopFlow framework. Implementing DFM or GTP individually significantly enhanced both A_{AUC} and A_{Last}, validating their respective contributions. DFM stabilizes learning by distilling domain-invariant self-knowledge and enhancing class discrimination in later layers. GTP preserves performance under varying batch compositions by consolidating transient batch cues into a persistent fea-

Table 7: The Effectiveness of TopFlow in Offline VIL Scenarios.

Method	iDigits		CORe50		DomainNet	
	Avg. Acc ↑	Forgetting ↓	Avg. Acc ↑	Forgetting ↓	Avg. Acc ↑	Forgetting ↓
ICON	70.67±2.13	13.49±2.09	79.61±1.73	8.19±1.47	49.08±2.11	17.88±2.40
TopFlow (Ours)	**73.33±1.72**	**12.14±4.20**	**79.77±1.47**	**8.03±1.19**	**49.62±1.82**	**17.30±3.09**

ture topology. Together, they provide complementary gains: DFM improves representations, while GTP maintains their structure for Online VIL.

Effectiveness of TopFlow in Standard CL Scenarios. As shown in Table 4, DFM and GTP improve performance in both standard CIL and DIL, individually and combined. We adopt CIFAR-100 [23] 10-split CIL and CORe50 [34] 8-split DIL as standard setups, with CODA-P [46] and S-Prompts [56] as baselines, respectively. Despite the limited domain variation in CIL, both components provide consistent gains, and the improvements are more pronounced in DIL, supporting their motivation under domain shifts.

Layer Selection for DFM. To analyze the effect of layer selection for Eq. 5, we conduct an ablation study using various (n, l) pairs in the ViT encoder (Table 5). Applying DFM to early layers such as (0,5) degrades performance, suggesting that low-level features are not well-aligned with the high-level semantics captured in the final layer. In contrast, using deeper layers such as (5,10) or (6,11) improves performance; combining (5,10) and (6,11) reaches 53.49, and our final configuration (6,11) achieves the best A_{Last} of 54.82. This indicates that later layers better preserve semantic information suitable for matching with final representations, yielding a favorable trade-off between abstraction and compatibility.

Table 6: Performance comparison of different methods and variants.

Method	A_{Last}
CODA-P	49.13
+ DFM	52.88
+ GTP	53.17
+ DFM, GTP	**54.62**
PEC	45.99
+ DFM	46.64
+ GTP	48.19
+ DFM, GTP	**48.86**

DFM and GTP with Other Models. Since DFM and GTP are designed as general components, we further examine whether they remain effective when integrated with other CL algorithms. We conducted ablation experiments to verify whether the proposed DFM and GTP could also yield performance improvements on baseline CL algorithms other than the MVP. As shown in Table 6, DFM and GTP lead to performance improvements over the baseline, even independently. DFM and GTP robustly improve performance over the baseline on the CORe50 dataset and other existing continual learning algorithms used in the main experiments.

Effectiveness of TopFlow in Offline VIL Scenarios. As shown in Table 7, the proposed TopFlow is effective not only in standard Offline CL scenarios such as CIL and DIL, but also in the more challenging Offline VIL setting, where both class and domain distributions change simultaneously. TopFlow achieves this by explicitly maintaining structural consistency by enforcing topology-preserving and Domain-agnostic matching across tasks, allowing newly learned representations to align with the previously learned feature structure. As a result, TopFlow achieves

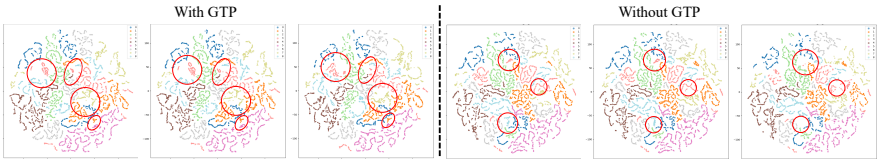


Fig. 4: t-SNE visualization of output features over tasks with GTP and without GTP.

stable representation learning and consistently improves performance across Offline VIL benchmarks. From a scenario perspective, TopFlow was originally designed for the more challenging *Online VIL* setting, where the model must adapt to non-stationary data streams without access to the full dataset distribution. Nevertheless, as shown in Table 7, TopFlow also achieves strong performance in the *Offline VIL* scenario. In contrast, the Offline VIL method ICON fails to maintain competitive performance in the Online VIL setting, as reported in Table 2. This result suggests that while methods tailored to Offline VIL may not generalize well to streaming environments, TopFlow provides a more robust solution that transfers across different VIL scenarios.

Visualization of GTP. To qualitatively analyze the effect of GTP, we visualize the output features using t-SNE [36] over tasks with and without GTP in Figure 4. We tracked the evolution of features from the first 10 classes, which are available for all tasks in the CORe50 dataset. And low-dimensional approximation is performed across all tasks to maintain consistency and relative positions of features. While it is challenging to project high-dimensional topological structures into 2D space, we observe that GTP helps to preserve the relative arrangement and internal structure of clusters across tasks. Especially in red circles, the meaningful topologies (e.g., relative connections or penetrations between clusters) are better maintained with GTP, whereas they drift and dismorph without it. This qualitative analysis supports our findings, demonstrating that GTP preserves global topological relationships in the feature space during online learning.

5 Conclusion

We proposed a new online continual learning scenario named Online VIL, which simulates a complex real world where states are ever-changing and there are no concepts of tasks and clear boundaries between them. Through analysis, we determined the direction for problem-solving in Online VIL and defined novel TopFlow framework. We demonstrated that the proposed TopFlow showed SOTA performance in the challenging Online VIL scenario, and its effectiveness through various experiments. Online VIL involves stochastic task construction, where the composition of data to each task is influenced by random seed. While this design captures more realistic dynamics, it can introduce variability in results. Despite this, we hope that our Online VIL scenario will serve as a new benchmark for advancing real-world incremental learning research, providing a more realistic and challenging setting for future studies.

Acknowledgements

This work was partly supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.RS-2026-25507543, Development of AI Co-Scientist based on Scientific Causal World Model, 80%), and in part by Korea Planning & Evaluation Institute of Industrial Technology (KEIT) grant funded by the Korea government (MOTIE) (RS-2024-00444344), and by the "Advanced GPU Utilization Support Program" funded by the Government of the Republic of Korea (Ministry of Science and ICT).

References

1. Aljundi, R., Belilovsky, E., Tuytelaars, T., Charlin, L., Caccia, M., Lin, M., Page-Caccia, L.: Online continual learning with maximal interfered retrieval. Advances in neural information processing systems **32** (2019)
2. Bang, J., Kim, H., Yoo, Y., Ha, J.W., Choi, J.: Rainbow memory: Continual learning with a memory of diverse samples. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8218–8227 (2021)
3. Cayton, L., et al.: Algorithms for manifold learning. Univ. of California at San Diego Tech. Rep **12**(1-17), 1 (2005)
4. Chen, D., Wang, D., Darrell, T., Ebrahimi, S.: Contrastive test-time adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 295–305 (2022)
5. De Lange, M., Aljundi, R., Masana, M., Parisot, S., Jia, X., Leonardis, A., Slabaugh, G., Tuytelaars, T.: A continual learning survey: Defying forgetting in classification tasks. IEEE transactions on pattern analysis and machine intelligence **44**(7), 3366–3385 (2021)
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 248–255. Ieee (2009)
7. Döbler, M., Marsden, R.A., Yang, B.: Robust mean teacher for continual and gradual test-time adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7704–7714 (2023)
8. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. In: Proceedings of the International Conference on Learning Representations (2020)
9. Ester, M., Kriegel, H.P., Sander, J., Xu, X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: kdd. vol. 96, pp. 226–231 (1996)
10. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: Proceedings of the International Conference on Machine Learning. pp. 1180–1189. PMLR (2015)
11. Gao, Q., Zhao, C., Sun, Y., Xi, T., Zhang, G., Ghanem, B., Zhang, J.: A unified continual learning framework with general parameter-efficient tuning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 11483–11493 (2023)

12. Gong, B., Shi, Y., Sha, F., Grauman, K.: Geodesic flow kernel for unsupervised domain adaptation. In: 2012 IEEE conference on computer vision and pattern recognition. pp. 2066–2073. IEEE (2012)
13. Gong, T., Jeong, J., Kim, T., Kim, Y., Shin, J., Lee, S.J.: Note: Robust continual test-time adaptation against temporal correlation. Advances in Neural Information Processing Systems **35**, 27253–27266 (2022)
14. Gopalan, R., Li, R., Chellappa, R.: Domain adaptation for object recognition: An unsupervised approach. In: 2011 international conference on computer vision. pp. 999–1006. IEEE (2011)
15. Gunasekara, N., Pfahringer, B., Gomes, H.M., Bifet, A.: Survey on online streaming continual learning. In: Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence. pp. 6628–6637 (2023)
16. Guo, Y., Liu, B., Zhao, D.: Online continual learning through mutual information maximization. In: International conference on machine learning. pp. 8109–8126. PMLR (2022)
17. Hartigan, J.A., Wong, M.A.: Algorithm as 136: A k-means clustering algorithm. Journal of the royal statistical society. series c (applied statistics) **28**(1), 100–108 (1979)
18. He, Y., Chen, Y., Jin, Y., Dong, S., Wei, X., Gong, Y.: Dyson: Dynamic feature space self-organization for online task-free class incremental learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 23741–23751 (2024)
19. Hendrycks, D., Basart, S., Mu, N., Kadavath, S., Wang, F., Dorundo, E., Desai, R., Zhu, T., Parajuli, S., Guo, M., et al.: The many faces of robustness: A critical analysis of out-of-distribution generalization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8340–8349 (2021)
20. Iwasawa, Y., Matsuo, Y.: Test-time classifier adjustment module for model-agnostic domain generalization. Advances in Neural Information Processing Systems **34**, 2427–2440 (2021)
21. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A.A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., et al.: Overcoming catastrophic forgetting in neural networks. Proceedings of the national academy of sciences **114**(13), 3521–3526 (2017)
22. Koh, H., Kim, D., Ha, J.W., Choi, J.: Online continual learning on class incremental blurry task configuration with anytime inference. arXiv preprint arXiv:2110.10031 (2021)
23. Krizhevsky, A., Hinton, G., et al.: Learning multiple layers of features from tiny images.(2009) (2009)
24. Kuhn, H.W.: The hungarian method for the assignment problem. Naval research logistics quarterly **2**(1-2), 83–97 (1955)
25. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (1998)
26. Lee, D., Yoon, J., Hwang, S.J.: Becotta: Input-dependent online blending of experts for continual test-time adaptation. arXiv preprint arXiv:2402.08712 (2024)
27. Li, D., Yang, Y., Song, Y.Z., Hospedales, T.M.: Deeper, broader and artier domain generalization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5542–5550 (2017)
28. Li, Y., Wang, N., Shi, J., Liu, J., Hou, X.: Revisiting batch normalization for practical domain adaptation. arXiv preprint arXiv:1603.04779 (2016)

29. Li, Z., Hoiem, D.: Learning without forgetting. IEEE Trans. Pattern Anal. Mach. Intell. **40** (2017). https://doi.org/10.1109/TPAMI.2017.2773081, https://doi.org/10.1109/TPAMI.2017.2773081
30. Li, Z., Hoiem, D.: Learning without forgetting. IEEE transactions on pattern analysis and machine intelligence **40**(12), 2935–2947 (2017)
31. Lin, Z., Shi, J., Pathak, D., Ramanan, D.: The clear benchmark: Continual learning on real-world imagery. In: Thirty-fifth conference on neural information processing systems datasets and benchmarks track (round 2) (2021)
32. Liu, S., Yang, Y., Li, X., Clifton, D.A., Ghanem, B.: Enhancing online continual learning with plug-and-play state space model and class-conditional mixture of discretization. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 20502–20511 (2025)
33. Liu, Y., Hong, X., Tao, X., Dong, S., Shi, J., Gong, Y.: Model behavior preserving for class-incremental learning. IEEE Transactions on Neural Networks and Learning Systems **34**(10), 7529–7540 (2022)
34. Lomonaco, V., Maltoni, D.: Core50: a new dataset and benchmark for continuous object recognition. In: Conference on Robot Learning. pp. 17–26. PMLR (2017)
35. Long, M., Cao, Y., Wang, J., Jordan, M.: Learning transferable features with deep adaptation networks. In: International conference on machine learning. pp. 97–105. PMLR (2015)
36. van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of Machine Learning Research **9**(86), 2579–2605 (2008), http://jmlr.org/papers/v9/vandermaaten08a.html
37. Mai, Z., Li, R., Kim, H., Sanner, S.: Supervised contrastive replay: Revisiting the nearest class mean classifier in online class-incremental continual learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3589–3599 (2021)
38. Moon, J.Y., Park, K.H., Kim, J.U., Park, G.M.: Online class incremental learning on stochastic blurry task boundary via mask and visual prompt tuning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 11731–11741 (2023)
39. Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., Ng, A.Y.: Reading digits in natural images with unsupervised feature learning. NeurIPS Workshop on Deep Learning and Unsupervised Feature Learning (2011)
40. Pan, X., Luo, P., Shi, J., Tang, X.: Two at once: Enhancing learning and generalization capacities via ibn-net. In: Proceedings of the european conference on computer vision (ECCV). pp. 464–479 (2018)
41. Park, K.H., Song, K., Park, G.M.: Pre-trained vision and language transformers are few-shot incremental learners. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 23881–23890 (2024)
42. Park, M.Y., Lee, J.H., Park, G.M.: Versatile incremental learning: Towards class and domain-agnostic incremental learning. In: European Conference on Computer Vision. pp. 271–288. Springer (2024)
43. Peng, X., Bai, Q., Xia, X., Huang, Z., Saenko, K., Wang, B.: Moment matching for multi-source domain adaptation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1406–1415 (2019)
44. Rolnick, D., Ahuja, A., Schwarz, J., Lillicrap, T., Wayne, G.: Experience replay for continual learning. Advances in neural information processing systems **32** (2019)
45. Sarfraz, S., Sharma, V., Stiefelhagen, R.: Efficient parameter-free clustering using first neighbor relations. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8934–8943 (2019)

46. Smith, J.S., Karlinsky, L., Gutta, V., Cascante-Bonilla, P., Kim, D., Arbelle, A., Panda, R., Feris, R., Kira, Z.: Coda-prompt: Continual decomposed attention-based prompting for rehearsal-free continual learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11909–11919 (2023)
47. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. Advances in neural information processing systems **30** (2017)
48. Song, J., Park, K., Shin, I., Woo, S., Zhang, C., Kweon, I.S.: Test-time adaptation in the dynamic world with compound domain knowledge management. IEEE Robotics and Automation Letters **8**(11), 7583–7590 (2023)
49. Tao, X., Hong, X., Chang, X., Dong, S., Wei, X., Gong, Y.: Few-shot class-incremental learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12183–12192 (2020)
50. Tao, X., Hong, X., Chang, X., Gong, Y.: Bi-objective continual learning: Learning 'new'while consolidating 'known'. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 5989–5996 (2020)
51. Torralba, A., Efros, A.A.: Unbiased look at dataset bias. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (2011)
52. Venkateswara, H., Eusebio, J., Chakraborty, S., Panchanathan, S.: Deep hashing network for unsupervised domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5018–5027 (2017)
53. Volpi, R., Larlus, D., Rogez, G.: Continual adaptation of visual representations via domain randomization and meta-learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4443–4453 (2021)
54. Wang, Q., Wang, R., Wu, Y., Jia, X., Meng, D.: Cba: Improving online continual learning via continual bias adaptor. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 19082–19092 (2023)
55. Wang, S., Shi, W., Dong, S., Gao, X., Song, X., Gong, Y.: Semantic knowledge guided class-incremental learning. IEEE Transactions on Circuits and Systems for Video Technology **33**(10), 5921–5931 (2023)
56. Wang, Y., Huang, Z., Hong, X.: S-prompts learning with pre-trained transformers: An occam's razor for domain incremental learning. In: Oh, A.H., Agarwal, A., Belgrave, D., Cho, K. (eds.) Proceedings of the Advances in Neural Information Processing Systems (2022)
57. Wang, Z., Zhang, Z., Ebrahimi, S., Sun, R., Zhang, H., Lee, C.Y., Ren, X., Su, G., Perot, V., Dy, J., et al.: Dualprompt: Complementary prompting for rehearsal-free continual learning. In: European Conference on Computer Vision. pp. 631–648. Springer (2022)
58. Wang, Z., Zhang, Z., Lee, C.Y., Zhang, H., Sun, R., Ren, X., Su, G., Perot, V., Dy, J., Pfister, T.: Learning to prompt for continual learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 139–149 (2022)
59. Wei, Y., Ye, J., Huang, Z., Zhang, J., Shan, H.: Online prototype learning for online continual learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 18764–18774 (2023)
60. Zając, M., Tuytelaars, T., van de Ven, G.M.: Prediction error-based classification for class-incremental learning. In: International Conference on Learning Representations (2024)
61. Zhang, G., Wang, L., Kang, G., Chen, L., Wei, Y.: Slca: Slow learner with classifier alignment for continual learning on a pre-trained model. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 19148–19158 (2023)

62. Zhang, Y., Mehra, A., Niu, S., Hamm, J.: Dpcore: Dynamic prompt coreset for continual test-time adaptation. arXiv preprint arXiv:2406.10737 (2024)
63. Zhou, D.W., Cai, Z.W., Ye, H.J., Zhang, L., Zhan, D.C.: Dual consolidation for pre-trained model-based domain-incremental learning. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 20547–20557 (2025)