

Forget Less by Learning from Parents Through Hierarchical Relationships

Arjun Ramesh Kaushik, Naresh Kumar Devulapally, Vishnu Suresh Lokhande, Nalini K. Ratha, Venu Govindaraju

University at Buffalo-SUNY
{kaushik3, devulapa, vishnulo, nratha, govind}@buffalo.edu

Abstract

Custom Diffusion Models (CDMs) offer impressive capabilities for personalization in generative modeling, yet they remain vulnerable to catastrophic forgetting when learning new concepts sequentially. Existing approaches primarily focus on minimizing interference between concepts, often neglecting the potential for positive inter-concept interactions. In this work, we present Forget Less by Learning from Parents (FLLP), a novel framework that introduces a parent-child inter-concept learning mechanism in hyperbolic space to mitigate forgetting. By embedding concept representations within a Lorentzian manifold, naturally suited to modeling tree-like hierarchies, we define parent-child relationships in which previously learned concepts serve as guidance for adapting to new ones. Our method not only preserves prior knowledge but also supports continual integration of new concepts. We validate FLLP on three public datasets and one synthetic benchmark, showing consistent improvements in both robustness and generalization.

Introduction

Custom diffusion models (CDMs) (Gal et al. 2022; Gu et al. 2023) have emerged as a powerful tool for personalizing image generation by adapting general-purpose text-to-image models to individual users or specific concepts. These models enable the synthesis of images from user-provided reference images, offering a pathway to generate personalized content. The growing interest in model customization reflects a broader goal of user-centric generative modeling, where models are tailored to represent distinct objects, styles, or identities that are not present in the training data. Text-conditioning enhances controllability and flexibility in diffusion models (Podell et al. 2024; Cheng et al. 2020) by guiding these models to generate semantically aligned visual content. Early personalization methods achieved this by fine-tuning a subset of model weights (Yu, Yang, and Zhang 2024) or by learning new token embeddings (Gal et al. 2022) from a handful of user images. While effective, these methods necessitate per-concept optimization, resulting in increased computational overhead and significant time requirements. Recent work addresses these limitations through tuning-free methods (Wang et al.

2025), without the need for re-training during inference. Beyond individual concept customization, recent research has focused on multi-concept personalization, composing multiple novel concepts into a single generation task (Kumari et al. 2023; Smith et al. 2024; Dong et al. 2024). Compared to single-concept settings, multi-concept customization entails diverse reference images and complex compositional prompts, leading to challenges such as handling previously unseen concepts, reducing interference among concepts, and maintaining prior knowledge across tasks. These challenges align naturally with the goals of continual learning, where a model learns incrementally from a sequence of concept-specific tasks without catastrophic forgetting. Recent mitigation strategies for catastrophic forgetting in CDMs, such as elastic weight consolidation, latent replay layers, and gradient matching techniques, primarily focus on isolating concepts through regularization and memory optimization. While these methods reduce immediate interference, they often fail to harness the latent relationships between learned concepts (Guo and Jin 2025; Dong et al. 2024). Unlike current machine learning systems, human learning naturally excels at transferring knowledge across tasks, an ability rooted in our capacity to organize concepts into hierarchical structures. Research shows that we dynamically update them to support both long-term retention and generalization. Further, recent findings reveal that the brain itself is wired to accumulate and refine such hierarchical conceptual knowledge (Theves et al. 2021).

Motivated by these insights, we introduce **Forget Less by Learning from Parents (FLLP)**, the first framework to leverage inter-concept interactions positively, to the best of our knowledge. **FLLP** explicitly models hierarchical concept relationships to emulate human-like knowledge transfer. Like cognitive hierarchies, our approach projects concept embeddings into hyperbolic space (Anderson 2006), a natural fit for encoding hierarchical structures due to its intrinsic tree-like geometry (Desai et al. 2023; Sala et al. 2018). Within this manifold, we enforce hierarchical parent entailment, requiring new concepts to reside within the entailment cones of their predecessors. This constraint mirrors how humans anchor novel skills (e.g., motorcycle control) to foundational schemas (bicycle balance), preserving conceptual dependencies while enabling adaptive reuse (Theves et al. 2021).

Our work makes three core contributions: (1) FLLP leverages natural hierarchies formed by hyperbolic embeddings to model inter-concept interactions. (2) We propose a Hierarchical Parent Entailment Loss using the union-find algorithm (Doyle and Rivest 1976) to form a parent-chain that enforces new concepts to lie within the parent entailment cones. (3) We utilize this hierarchical learning to train CDMs continually and achieve SOTA results across three public datasets and one synthetic benchmark. To the best of our knowledge, this is the first contribution towards *leveraging* inter-concept interactions rather than mitigating them for a continual learning problem in CDMs.

Related Work

Incremental Concept Learning. Recent class-incremental learning methods focus on mitigating catastrophic forgetting via regularization, distillation, and rehearsal, but are largely developed for discriminative classifiers with fixed label spaces rather than generative models with continuous concept manifolds (Zhang et al. 2025). The survey in (Zhang et al. 2025) emphasizes that interference between related classes remains a central limitation even under strong regularization or memory mechanisms. In text-to-image diffusion, lifelong and personalized approaches such as L2DM (Sun et al. 2024), C-LoRA (Smith et al. 2024), and Concept-Guard (Guo and Jin 2025) adapt these ideas to the noise-prediction space by attaching concept-aware adapters and distillation losses to preserve old concepts during continual customization. However, they still treat each concept as an essentially independent unit, offering only coarse control over intra-concept variation and higher-order concept relationships.

Intra-Concept Learning Dynamics. Multi-concept customization work shows that concept-specific updates interact in highly non-linear ways. Multi-Concept Customization, Mix-of-Show, and Multi-LoRA all report that composing several learned concepts or LoRA adapters can lead to strong interference and style leakage, especially when concepts are visually similar (Kumari et al. 2023; Gu et al. 2023; Zhong et al. 2024). Cross-Attention Bootstrapping further reveals that cross-attention heads in diffusion models specialize to distinct object and attribute factors, and that misaligned heads are a key source of spurious edits and concept entanglement (Zhao et al. 2025). Together, these results motivate architectures that explicitly separate and recombine sub-concepts rather than treating each personalized token or adapter as monolithic.

Catastrophic Forgetting in Concept Diffusion. Within text-to-image diffusion specifically, forgetting and confusion arise because concept information is distributed across the UNet, cross-attention layers, and token embeddings and is refined across timesteps. L2DM and C-LoRA retain concepts by distilling the denoising trajectory of earlier models and routing gradients through concept-specific low-rank adapters (Sun et al. 2024; Smith et al. 2024), while Concept-Guard disentangles forgetting from confusion and shows that overlapping attention patterns between related concepts are a dominant failure mode (Guo and Jin 2025). These systems markedly reduce catastrophic forgetting, but their pa-

rameter cost grows roughly linearly with the number of concepts and they still provide only implicit control over inter- and intra-concept interactions.

Hyperbolic Geometry. Standard latent diffusion models operate in Euclidean latent spaces (Rombach et al. 2022), which are poorly suited for representing deep hierarchies and tree-like concept structure. Hyperbolic vision-language models such demonstrate that negative curvature and exponential volume growth enable compact, hierarchy-aware image-text embeddings with improved semantic organization (Desai et al. 2023). HypDiff (Fu et al. 2024) proposes a hyperbolic latent diffusion model for graph generation, so as to preserve topological properties such as hierarchy and community structure during generation. However, these methods largely treat hyperbolic geometry as a static representation choice and do not address continual adaptation or concept interference in text-to-image diffusion.

Limitations of Current Approaches. Across these lines of work, three gaps remain: (1) continual diffusion methods prioritize per-concept preservation but model concepts as independent, limiting intra-concept specialization and compositional recombination (Sun et al. 2024; Smith et al. 2024; Guo and Jin 2025); (2) multi-concept customization exposes strong interference between learned edits without providing a principled latent structure for resolving it (Kumari et al. 2023; Gu et al. 2023; Zhong et al. 2024); and (3) hyperbolic generative models exploit hierarchical geometry but are not designed for dynamic, concept-aware updates over time (Desai et al. 2023; Liu et al. 2023). FLLP targets this intersection by coupling concept-aware continual updates with an explicitly structured latent space that is designed to accommodate evolving concept hierarchies.

Problem Definition

We consider the problem setting introduced in CIFC (Dong et al. 2024), which considers a sequential inflow of concepts. The model learns from an undefined series of text-guided concept customization tasks $T = \{T_c\}_{c=1}^C$, where C denotes the total number of tasks. Each task T_c consists of a dataset $T_c = \{(x_c^k, p_c^k, Y_c)\}_{k=1}^N$ where N is the number of reference images in the task, x_c^k is an image, p_c^k is a text prompt (e.g., “photo of a [V_{cat}]”), and Y_c represents the concept space. The setting supports diverse customization tasks, including multi-concept generation (Zhong et al. 2024), style transfer (Zhang et al. 2023), and image editing (Chen et al. 2024). There are three constraints in this problem - (1) **All concepts are distinct:** $Y_c \cap \left(\bigcup_{i=1}^{c-1} Y_i\right) = \emptyset$ indicating that new concepts in task c are distinct from those learned in the other $C - 1$ tasks. (2) **Concepts are learned in the order of inflow:** $T = \{T_1, T_2, \dots, T_C\}$, where the c -th concept T_c is learned with access only to prior concepts $\{T_1, \dots, T_{c-1}\}$ for $c = 1, \dots, C$. This models a unidirectional learning flow, preventing access to future unseen concepts. (3) **No-replay constraint:** No memory storage is allocated to retain training images from past tasks, ensuring purely incremental learning of personalized concepts. This setting allows models to continuously adapt to new personalization tasks while being

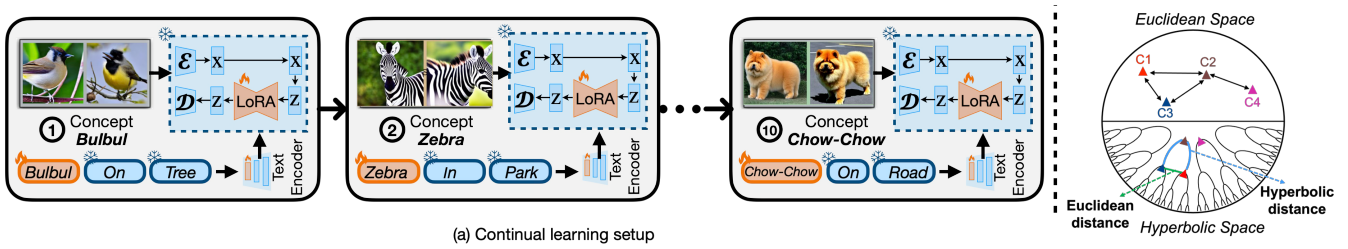


Figure 1: **Continual Learning for Concept Personalization and Hyperbolic Geometry.** (a) The Continual Learning (CL) paradigm involves training a model on a sequence of tasks or concepts (e.g., Bulbul, Zebra, Chow-chow) over time. A major challenge in this setup is Catastrophic Forgetting, where the model tends to lose knowledge of previously learned concepts as it acquires new ones. (b) In Euclidean space, all concepts coexist without any inherent hierarchy. In contrast, hyperbolic geometry enables hierarchical representation of concepts. Importantly, hyperbolic distances (blue line) are fundamentally different from Euclidean distances (green line).

exposed to catastrophic forgetting of prior concepts.

Understanding Catastrophic Forgetting

Catastrophic forgetting refers to the tendency of neural networks to lose previously acquired knowledge when trained on new tasks. This phenomenon arises from the stability-plasticity dilemma, where neural networks must balance retaining existing information with adapting to new inputs (Chen and Liu 2018). Catastrophic forgetting occurs when parameter updates for a new task T_{new} overwrite critical weight configurations optimized for prior tasks T_{old} : $P_{\text{old}}(W) \ll P_{\text{new}}(W)$, where W represents network weights, and P denotes task performance. This work advances the central hypothesis that catastrophic forgetting can be fundamentally mitigated through the use of hierarchical knowledge transfer mechanisms.

1D Gaussian Setup

To enable a systematic and interpretable analysis of catastrophic forgetting, we construct a synthetic one-dimensional dataset that serves as a controlled surrogate benchmark (see fig. 2). This benchmark comprises five sequential tasks, each defined by a Gaussian distribution with a distinct mean and unit variance, thereby emulating the concept drift characteristic of continual learning scenarios. A shared 1D UNet diffusion model is employed to reconstruct the data distribution associated with each task (see appendix for architectural details). To quantify forgetting, we compute the sum of absolute differences between the generated and true means across tasks. This serves as our primary evaluation metric, referred to as the *forgetting rate*, with lower values indicating improved retention of previously learned concepts.

What Causes Catastrophic Forgetting?

The catastrophic forgetting observed in the 1D Gaussian benchmark fig. 2 stems from two interrelated mechanisms: *parameter interference* and *task-specific representation collapse*. When training sequentially on tasks μ_i , gradient updates for new tasks: $\Delta W \propto \nabla_W \mathcal{L}_{\text{MSE}}(\mu_{\text{new}})$, perturb the network’s shared parameters W , overwriting features critical for reconstructing prior distributions μ_{old} . This interfer-

ence is amplified by the UNet’s frozen architecture, which forces all tasks to compete for representation space through a fixed set of weights.

Parameter interference emerges from fundamental limitations in gradient-based optimization. When learning task μ_{new} , the model computes weight updates proportional to the gradient of the current task’s MSE loss: $\Delta W \propto -\eta \frac{\partial}{\partial W} \|Gen(\mu_{\text{new}}) - \mu_{\text{new}}\|^2$ where Gen represents the generated output. These updates create vector fields in parameter space that optimally reduce μ_{new} ’s reconstruction error but catastrophically displace solutions for previous tasks μ_{old} . In our 1D setup, the frozen UNet architecture compounds this issue; though its core weights remain fixed, the task-specific token embeddings must encode all distributional shifts through a single projection layer. This creates a representational bottleneck where embedding vectors compete to map multiple Gaussian means (μ_1, \dots, μ_5) through identical network transformations, leading to destructive interference patterns (Lokhande et al. 2022).

Task-specific representation collapse occurs when sequential training causes embeddings for distinct μ_i to converge toward a shared subspace. The model’s token embeddings $\{E(\mu_i)\}$ gradually align with the dominant gradient directions from new tasks at t -th training step: $E(\mu_i)^{t+1} = E(\mu_i)^t - \eta \sum_{j=1}^k \frac{\partial \mathcal{L}_{\text{MSE}}}{\partial E(\mu_j)}$. With no task-specific signals, the model optimizes for a single set of parameters to minimize the aggregate reconstruction error across all tasks. The mean of means emerges as the Nash equilibrium for minimizing the reconstruction error. This is demonstrated in the first row of Fig. 2, where the model converges near the mean-of-means 9.6, and achieves a *forgetting rate* of 18.6.

Analysis: Fig. 2 illustrates the visual outcome of our experiments. Each subplot compares the true and generated distributions for a given task after training on the final task, denoted in blue and red colors, respectively. The matrix layout of subplots is sorted in descending order of *forgetting rate* in different approaches. From top to bottom, we observe a monotonic reduction in error, indicating that the model increasingly preserves information about earlier concepts. The visual alignment of blue (true) and red (generated) bumps in lower rows highlights effective knowledge retention, a de-

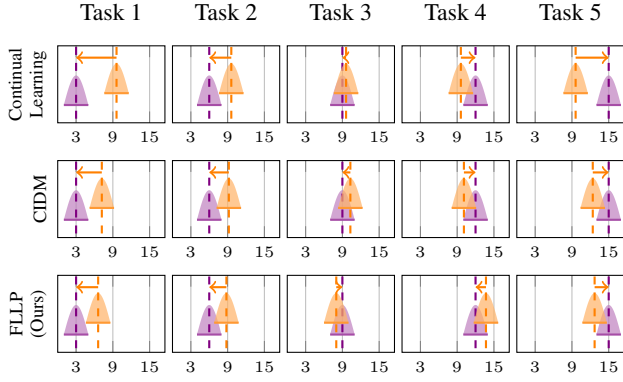


Figure 2: **Quantitative Analysis of 1D Gaussian.** The length of the orange arrows reflects the *forgetting rate* for each concept, with shorter arrows indicating better retention. CIDM (Dong et al. 2024) improves upon the baseline by achieving a lower *forgetting rate* of 13.2. FLLP (Ours) further reduces this to 11.4. The visibly shorter orange arrows in the bottom row corroborate the improved knowledge retention of FLLP over CIDM. (Best viewed by zooming in)

sirable trait for continual generative modeling. Notably, this alignment is most evident in Task 1, Task 2, and Task 3, where the extent of forgetting is also visualized by the length of the arrow indicating the mean shift.

While we have not yet formally introduced our proposed method, FLLP, we offer a brief preview to anchor the reader’s attention. CIDM (Dong et al. 2024) achieves a forgetting rate of 13.2, whereas our approach, FLLP, lowers this further to 11.4. For context, the vanilla continual learning baseline yields a significantly higher forgetting rate of 18.6. As a reminder, lower forgetting rates indicate better retention of previously learned concepts.

Hyperbolic Trees Enable Optimal Grouping of Concepts

In this section, we discuss the learning objectives of FLLP to model inter-concept interactions through hyperbolic representations of previously learned concepts and incoming concepts. We begin by introducing hyperbolic geometry and the tools required to implement our methodology. To adapt the c -th customization task T_c , we employ LoRA (Hu et al. 2022; Wang et al. 2023) to fine-tune the pretrained denoising UNet, commonly denoted using ϵ .

Background into Hyperbolic Geometry

We adopt the Lorentz model of hyperbolic geometry for developing our framework. In this model, the n -dimensional hyperbolic space \mathbb{H}^n is realized as the upper sheet of a two-sheeted hyperboloid embedded in \mathbb{R}^{n+1} , endowed with the Lorentzian inner product (See Sec.). The model’s key advantage lies in its ability to express geodesics, distances, and projections in closed form, which is highly beneficial for optimization in high-dimensional settings. Following the study of special relativity theory (Lorentz et al. 1923), the

axis of symmetry of the hyperboloid is referred to as the *time dimension*, while the remaining n axes form the *space dimensions*. We adopt this terminology throughout, writing vectors in \mathbb{R}^{n+1} as $[\mathbf{x}_{\text{space}}, x_{\text{time}}]$. We will now discuss a few tools from hyperbolic geometry to realize the learning objectives for our model.

Geodesics: The shortest path between two points on a Riemannian manifold is defined as a *geodesic*. In the Lorentz model of hyperbolic space, geodesics are defined as the intersections between the hyperboloid and hyperplanes that pass through the origin of the ambient space \mathbb{R}^{n+1} . The Lorentzian distance between two points $x, y \in \mathcal{L}^n$, where k is the constant negative curvature, is given by $d_{\mathcal{L}}(x, y) = \frac{1}{\sqrt{k}} \cosh^{-1}(-k\langle x, y \rangle_{\mathcal{L}})$.

Aperture: The *aperture* defines the half-angle of the entailment cone originating from a point x in hyperbolic space, for a small constant K (typically set to 0.1) is defined as $\text{Aper}(x) = \sin^{-1}\left(\frac{2K\sqrt{k}}{\|\mathbf{x}_{\text{space}}\|}\right)$.

Exterior Angle: The *exterior angle* between two Lorentzian embeddings x and y measures how far y lies from the central axis of x ’s entailment cone. This is defined as the angle between x and y relative to the origin O , subtracted from π ,

$$\text{Ext}(x, y) = \cos^{-1}\left(\frac{y_{\text{time}} + x_{\text{time}}k\langle x, y \rangle_{\mathcal{L}}}{\|\mathbf{x}_{\text{space}}\| \sqrt{(k\langle x, y \rangle_{\mathcal{L}})^2 - 1}}\right)$$

Continual Learning in Diffusion Models using Hyperbolic Geometry

We present a novel learning objective - **Hyperbolic Parent Entailment Loss** - to model hierarchical relationships between concepts using hyperbolic geometry in Algorithm 1 (Appendix). This loss function is governed by three crucial design choices: (1) transferring embeddings from Euclidean space to Lorentz hyperboloid, (2) finding the parent concept(s) for the incoming concept, and (3) designing a training objective to ensure the child concept (incoming concept) is learned through parent guidance (previously learned concepts).

Transferring embeddings to hyperboloid: Let an embedding vector be $\mathbf{v} \in \mathbb{R}^n$. We aim to apply a transformation such that the resulting vector \mathbf{x} lies on the Lorentz hyperboloid $\mathcal{L}^n \subset \mathbb{R}^{n+1}$. Let the vector $\mathbf{v} = [\mathbf{v}, 0] \in \mathbb{R}^{n+1}$. We observe that \mathbf{v} belongs to the tangent space at the origin \mathcal{O} of the hyperboloid, since the Lorentzian inner product satisfies: $\langle \mathcal{O}, \mathbf{v} \rangle_{\mathcal{L}} = 0$. Thus, we parameterize only the space components of the Lorentz model (i.e., $\mathbf{v} = \mathbf{v}_{\text{space}}$). This parameterization simplifies the exponential map. Instead of computing the full map in \mathbb{R}^{n+1} , we write the exponential map

using only the space components: $\mathbf{x}_{\text{space}} = \frac{\sinh(\sqrt{c}\|\mathbf{v}_{\text{space}}\|)}{\sqrt{c}\|\mathbf{v}_{\text{space}}\|} \cdot \mathbf{v}_{\text{space}}$. Since the first term is zero, and the Lorentzian norm of \mathbf{v} simplifies to the Euclidean norm of the space components: $\|\mathbf{v}\|_{\mathcal{L}}^2 = \langle \mathbf{v}, \mathbf{v} \rangle_{\mathcal{L}} = \langle \mathbf{v}_{\text{space}}, \mathbf{v}_{\text{space}} \rangle = \|\mathbf{v}_{\text{space}}\|^2$. The corresponding time component x_{time} can then be computed from $\mathbf{x}_{\text{space}}$ as shown in Sec. . The resulting vector $\mathbf{x} = [x_{\text{time}}, \mathbf{x}_{\text{space}}]$ lies on the Lorentz hyperboloid, eliminating the need for an orthogonal projection step. This parameterization is similar to (Desai et al. 2023) and simpler than

prior approaches that operate in the full ambient space \mathbb{R}^{n+1} (Law et al. 2019; Le et al. 2019).

Parent Concept Identification: Let the CLIP (Ramesh et al. 2022) image features of the c -th concept be $v_c \in \mathbb{R}^{s \times n}$, where s is the number of source images and n is the feature dimensionality. Let $t_c \in \mathbb{R}^{(c-1) \times \text{len}(Y_c) \times n}$ represent the learned token embeddings corresponding to $c-1$ previously known concepts, where $\text{len}(Y_c)$ corresponds to the number of tokens. The image features and learned token embeddings are averaged to obtain $\bar{v}_c \in \mathbb{R}^{1 \times n}$ and $\bar{t}_c \in \mathbb{R}^{(c-1) \times n}$.

We define a recursive search strategy to construct a parent chain for the new concept, denoted by \mathcal{P}_{new} , where each parent will serve as an anchor. The intuitive idea is to iteratively move up the tree from the leaf node to root-level abstractions, as detailed in the Algorithm. 2 (Appendix).

Let Z_{img} and Z_{old} denote the hyperbolic projections of \bar{v}_c and \bar{t}_c , respectively. We initialize the current embedding $\mathbf{z}_{\text{curr}} \leftarrow Z_{\text{img}}$ and construct a search pool $\mathcal{S} = Z_{\text{old}} \cup \{Z_{\text{img}}\}$, indexed by $\mathcal{I} = \{0, \dots, c-1\} \cup \{-1\}$, where -1 refers to the c -th concept. At each iteration, we compute the negative Lorentzian geodesic distance between \mathbf{z}_{curr} and all embeddings in \mathcal{S} , selecting the second-closest index (skipping the trivial self-match) as the parent candidate. If the parent index is already visited or corresponds to the new concept itself, we detect a loop and terminate. Otherwise, the selected parent is assigned to \mathcal{P}_{new} , and \mathbf{z}_{curr} is updated accordingly. This process recursively builds a chain of semantically relevant parent concepts, revealing the new concept’s hierarchical alignment with seen concepts in hyperbolic space. Note that the Lorentzian geodesic distance $d_{\mathcal{L}}$ captures the exponential separation of nodes from the origin.

Storage of Image Attention Maps: For each of the $c-1$ previously observed image attention maps, denoted by I_α for $\alpha \in \{1, \dots, c-1\}$, we store their timestep-weighted mean representation. This is computed as: $\bar{I}_\alpha = \sum_{ts=1}^{\text{timesteps}} \frac{1}{ts} I_\alpha^{(ts)}$, where $I_\alpha^{(ts)}$ denotes the attention map at timestep ts . This aggregation provides a more stable and representative summary of each concept’s attention behavior over time. Note that this does not violate the “no memory” constraint, as the constraint is aimed at restricting the storage of reference images of previous concepts.

Inducing Hierarchical Structure Between Parent-Child Concepts: In hyperbolic space, an *entailment cone* C_u at a point u represents the region formed by a hyperbolic cone (Ganea, Becigneul, and Hofmann 2018; Le et al. 2019). These cones narrow as the point u moves farther from the origin. The half-aperture of C_u is defined in Eq. . Let x and y represent the hyperbolic projections (spatial and time components) of the image attention maps corresponding to a child and parent concept, respectively. The associated time components are computed separately, as shown in Sec. . Inspired by prior work on hierarchical embeddings (Ganea, Becigneul, and Hofmann 2018), we utilize entailment cones to enforce a partial ordering between the image attention maps of concepts. In particular, we constrain the child concept x to lie within the entailment cone defined by its parent y . To encourage this behavior during training, we introduce a

loss term that penalizes violations of this constraint. Specifically, we impose a penalty when the exterior angle between x and y exceeds the aperture of y , and β is a hyperparameter that controls the flexibility of the constraint:

$$\mathcal{L}_{\text{entailParent}} = \max(0, \text{ext}(y, x) - \beta * \text{aper}(y)) \quad (1)$$

Training

Following CIDM (Dong et al. 2024), we capture semantic patterns that persist in different tasks through direct supervision of LoRA weights. Specifically, a layer-wise common subspace \mathbf{W}_*^l is introduced to share across tasks. Additionally, a learnable projection matrix \mathbf{H}_i^l encodes common semantic attributes. Together, we optimize by minimizing the reconstruction loss: $\mathcal{L}_1 = \sum_{i=1}^g \sum_{l=1}^L \|\Delta \mathbf{W}_i^l - \mathbf{H}_i^l \mathbf{W}_*^l\|_F^2$. Finally, we modify the total loss function for the Custom Diffusion model (Refer Sec. in appendix) with $\mathbf{m}_c = \{\mathbf{m}_c^l\}_{l=1}^L$ representing layer-wise textual embeddings, and hyperparameter γ . Our proposed framework, FLLP, is trained on $\mathcal{L}_{\text{total}}$ utilizing a dual-step optimization strategy (Dong et al. 2024).

$$\mathcal{L}_{\text{total}} = \mathbb{E}_{\substack{z \sim \Phi(x_c^k), m_c \\ \epsilon \sim \mathcal{N}(0, I), t}} \left[\|\epsilon - \epsilon_{\theta'}(z | m_c, t)\|_2^2 + \gamma_1 \mathcal{L}_{\text{entailParent}} + \gamma_2 \mathcal{L}_1 \right] \quad (2)$$

Experiments

To rigorously demonstrate the robustness and generalizability of our proposed model, we conduct comprehensive experiments across three diverse public datasets, evaluating performance using CLIP (Ramesh et al. 2022) Image Alignment (IA) and Text Alignment (TA) metrics. Each class is treated as a distinct concept in our continual learning setup. Instead of leveraging all available classes, we intentionally select a subset of ten to simulate the sequential arrival of new concepts, an essential characteristic of realistic continual learning. This controlled design is critical for isolating and faithfully evaluating the effects of catastrophic forgetting under meaningful, yet analytically tractable, conditions. **Remark (Scope of Concept Evaluation):** In this work, we focus our analysis on a fixed set of 10 concepts across all experiments. This choice is informed by the architectural constraint imposed by the CLIP tokenizer, which allows only 77 learnable tokens (Ramesh et al. 2022). Since each concept typically consumes 2–3 tokens, the upper bound for feasible training is approximately 35 concepts. While we do present an ablation study extending beyond 10 concepts to examine scalability, our core evaluation remains grounded in 10 concepts, a practical and representative regime for studying catastrophic forgetting. Importantly, our methodology does not inherently limit extension to larger concept sets. Implementation and dataset details are provided in the Appendix.

Quantitative and Qualitative Analysis

CIFC Dataset: As shown in Tab. 1, we achieve superior IA and TA scores compared to CIDM (Dong et al. 2024) in 9 out of 10 concepts. Notably, FLLP achieved +9.2 IA scores on the C7 (Dog2) and +3.2, being the largest TA gain on the C9 (Cat2). Over the entire dataset, FLLP demonstrates a gain in both IA and TA scores of +1.2 and 1.1,

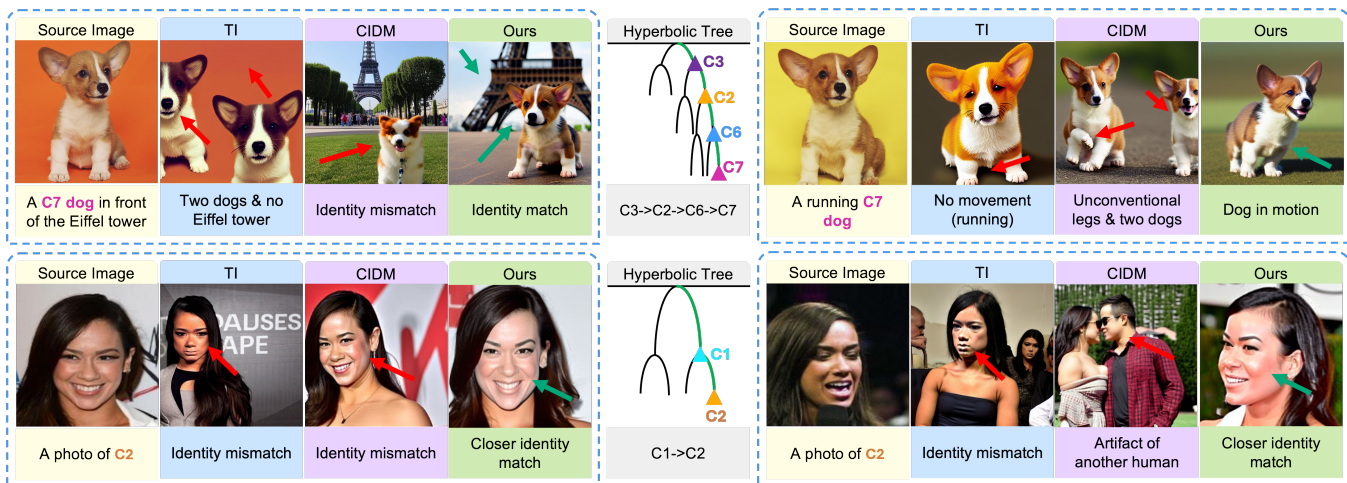


Figure 3: **Qualitative Analysis.** We compare the generated images in TI (Dosovitskiy et al. 2021), CIDM (Dong et al. 2024) and FLLP (Ours) on CelebA (Liu et al. 2015) and CIFC (Dong et al. 2024). The red and green arrows indicate regions of undesirable and desirable qualities, and their reasons are stated below each image. The hyperbolic tree indicates the parent chain of concepts that the model traversed in learning the incoming concept.

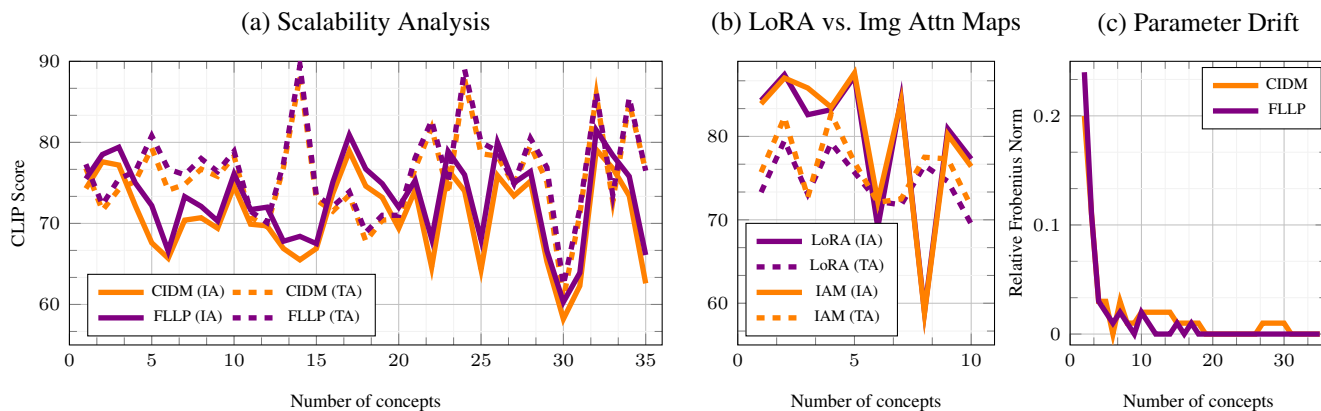


Figure 4: **Ablation Studies.** (a) FLLP scales to 35 concepts (limited by CLIP tokenizer) and consistently outperforms CIDM (Dong et al. 2024) on both IA and TA metrics. (b) Hyperbolic Parent Entailment Loss on LoRA weights benefits IA scores, but hurts TA scores. Thereby making Image Attention Maps (IAM) better suited for the loss function. (c) FLLP exhibits 22% lower parameter drift than CIDM, indicating the utilization of inter-concept interactions positively.

respectively. Note that each concept is trained with a different value of β (Threshold parameter). **CelebA Dataset:** On the CelebA dataset (Liu et al. 2015), FLLP surpasses CIDM with a notable gain of +4.4 in Image Alignment (IA) and +2.0 in Text Alignment (TA) scores. Among individual concepts, Person 5 (C5) exhibits the highest IA improvement with a margin of +8.9, while Person 8 (C8) shows the most significant TA boost, achieving +4.2 over CIDM. **ImageNet Dataset:** On the third benchmark dataset, FLLP continues to outperform CIDM, achieving gains of +1.1 in IA and +0.5 in TA scores. Performance improvements are observed consistently across most concepts, with C9 (Wood Rabbit) exhibiting the largest IA gain of +4.4. *FLLP consistently outperforms CIDM across CIFC, CelebA, and ImageNet, showing notable gains in both IA and TA metrics.*

Qualitative Analysis: We provide a qualitative comparison of the synthesized outputs generated by three approaches in Fig. 3: TI (Dosovitskiy et al. 2021), CIDM (Dong et al. 2024), and our proposed FLLP framework, spanning CelebA (Liu et al. 2015) (top row) and CIFC (Dong et al. 2024) (bottom row) datasets. Each image is generated using a source image along with its corresponding text prompt. Colored arrows are overlaid to highlight perceptual differences: green arrows denote regions where the output exhibits desirable qualities such as identity preservation or faithful prompt alignment; red arrows highlight deficiencies such as concept drift, artifact introduction, or semantic inconsistency. Commonly, we observe identity mismatches and semantic misalignment across both datasets in previous frameworks that FLLP corrects. Additionally, we visualize

Methods	C1		C2		C3		C4		C5		C6		C7		C8		C9		C10		Avg.		
	IA	TA	IA	TA	IA	TA	IA	TA	IA	TA	IA	TA	IA	TA	IA	TA	IA	TA	IA	TA	IA	TA	
CIFC dataset	Previous frameworks																						
	Finetuning	77.6	64.4	82.2	74.6	79.0	69.4	77.6	68.6	79.6	75.0	62.9	70.0	71.5	<u>76.7</u>	53.7	69.2	81.4	65.4	72.1	67.2	73.7	70.0
	EWC (Kirkpatrick et al. 2017)	78.7	67.1	83.8	77.5	80.4	72.7	80.3	77.9	80.7	76.7	64.0	<u>72.3</u>	76.5	74.2	57.1	72.0	84.4	66.0	73.1	70.4	75.9	72.7
	LWF (Li and Hoiem 2018)	80.4	70.8	79.7	75.2	80.9	71.0	77.4	77.4	80.9	76.0	61.8	71.7	73.2	76.3	53.5	72.9	78.1	72.5	74.7	70.0	74.1	73.4
	CLoRA (Smith et al. 2024)	83.2	69.4	83.4	78.0	81.1	<u>74.1</u>	80.6	78.8	84.9	76.4	66.3	69.6	76.2	<u>76.7</u>	58.1	73.9	83.0	69.0	72.1	<u>71.8</u>	76.9	73.6
	L2DM (Sun et al. 2024)	78.7	68.6	86.3	79.5	76.6	70.1	80.7	73.0	86.8	76.7	70.8	67.7	70.0	75.9	59.3	74.1	77.7	71.8	74.1	69.4	76.1	72.7
	TI (Dosovitskiy et al. 2021)	64.9	71.8	56.8	59.6	64.3	70.2	63.1	73.1	61.4	66.1	58.3	58.2	65.7	66.1	57.0	57.4	58.7	69.7	64.2	68.8	61.4	66.1
	Comparison against SOTA framework																						
	CIDM(Dong et al. 2024)	83.6	75.3	86.4	78.1	82.9	74.0	80.8	81.1	86.5	78.2	69.5	70.1	73.7	74.7	56.9	74.3	82.4	73.5	75.9	70.2	78.0	74.8
	FLLP (Ours)	83.9	<u>75.7</u>	87.0	<u>82.3</u>	85.8	72.7	83.5	<u>82.7</u>	87.6	<u>76.9</u>	72.2	72.1	84.2	72.5	58.8	<u>77.5</u>	80.2	<u>77.3</u>	76.4	71.6	80.0	<u>76.1</u>
Δ	+0.3	+0.4	+0.6	+4.2	+2.9	-1.3	+2.7	+1.6	+1.1	-1.3	+2.7	+2.0	+10.5	-2.2	+1.9	+3.2	-2.2	+3.8	+0.5	+1.4	+2.0	+1.3	
CelebA dataset	Previous frameworks																						
	TI (Dosovitskiy et al. 2021)	44.3	<u>62.4</u>	47.2	<u>65.2</u>	49.6	<u>61.5</u>	52.5	57.4	50.2	58.7	49.8	54.7	51.0	53.7	51.3	56.4	41.7	<u>61.9</u>	50.0	<u>65.3</u>	48.8	59.7
	Comparison against SOTA framework																						
	CIDM(Dong et al. 2024)	74.2	59.0	67.6	60.4	73.0	59.4	75.4	58.2	70.1	58.0	78.1	59.0	81.0	58.4	70.0	58.0	70.1	62.1	73.0	55.4	73.3	58.8
	FLLP (Ours)	77.4	58.9	74.7	61.7	73.4	60.8	78.7	<u>63.0</u>	79.0	<u>60.0</u>	82.4	<u>60.5</u>	84.7	<u>61.5</u>	75.8	<u>62.2</u>	75.4	61.7	75.6	57.7	77.7	<u>60.8</u>
Δ	+3.2	-0.1	+7.1	+1.3	+0.4	+1.4	+3.3	+4.8	+8.9	+2.0	+4.3	+1.5	+3.7	+3.1	+5.8	+4.2	+5.3	-0.4	+2.6	+2.3	+4.4	+2.0	
ImageNet dataset	Previous frameworks																						
	TI (Dosovitskiy et al. 2021)	62.2	73.4	66.7	76.5	62.0	62.4	64.0	67.2	64.1	66.1	67.3	<u>76.0</u>	63.1	63.6	61.8	<u>78.2</u>	65.8	76.5	66.2	72.7	64.4	71.3
	Comparison against SOTA framework																						
	CIDM(Dong et al. 2024)	80.8	<u>74.1</u>	71.4	77.9	81.3	81.7	83.0	82.6	79.2	83.7	84.5	69.6	76.7	<u>79.3</u>	87.5	73.7	82.2	<u>79.9</u>	85.0	<u>82.8</u>	81.2	78.5
	FLLP (Ours)	82.1	73.9	73.4	<u>78.9</u>	82.6	<u>82.5</u>	83.1	<u>84.0</u>	79.7	<u>85.1</u>	82.2	69.7	75.5	79.2	90.0	75.0	86.4	79.8	87.5	80.9	82.3	<u>79.0</u>
Δ	+1.3	-0.2	+2.0	+1.0	+1.3	+0.8	+0.1	+1.4	+0.5	+1.4	-2.3	+0.1	-1.2	-0.1	+2.5	+1.3	+4.4	-0.1	+2.5	-1.9	+1.1	+0.5	

Table 1: **Quantitative Analysis.** We compare our method **FLLP** against previous frameworks and the SOTA framework CIDM (Dong et al. 2024) across three datasets - CIFC (Dong et al. 2024), CelebA (Liu et al. 2015) and ImageNet (Deng et al. 2009). We observe an improvement in both Image Alignment (IA) and Text Alignment (TA) scores in all datasets. The higher IA score has been denoted in **bold**, and the higher TA value has been underlined.

the hyperbolic parent-child structure that FLLP leverages during training through a hyperbolic tree. These trees represent the conceptual trajectory taken by the model when incorporating a new concept. *For example, in the case of C7 dog (from the CIFC dataset), FLLP uses a hierarchical parent chain consisting of Cat1 (C3), Rubber Duck (C2), and Painting (C6) to guide the learning of the new concept.* In other words, concepts C2, C3, C6 and C7 form a group. Additional examples have been provided in the Appendix. *FLLP enhances output quality by correcting identity and semantic errors through hierarchical hyperbolic guidance.*

Ablation Studies

Why image attention maps instead of LoRA weights?

Direct hyperbolic optimization on LoRA weights forces a simultaneous geometric transformation of both modalities. Image-related weights benefit, while text-related weights suffer (Ramasinghe et al. 2024), resulting in a zero-sum game as elucidated in Fig. 4 (b) and Table 3 (Appendix).

On the effect of β : As illustrated in Fig. 6, 7, 8 (Appendix) across the three datasets, each concept exhibits a distinct optimal threshold value. The threshold modulates the extent of parental influence on child concepts; lower values enforce stronger guidance, while higher values relax this constraint.

Scalability: To evaluate FLLP beyond ten concepts, we conduct experiments on the CustomConcept101 dataset (Kumari et al. 2023). Due to CLIP’s tokenizer constraint allowing only 77 additional tokens (Ramesh et al. 2022), we are limited to training on 35 concepts (considering 2–3 tokens

per concept). As shown in Fig. 4(a) and Table 4 (Appendix), FLLP outperforms CIDM by +2.1 on average IA scores and +1.0 on average TA scores.

Comparing parameter drifts: We compare the evolution of LoRA weights in FLLP (Ours) and CIDM (Dong et al. 2024) using the Frobenius norm averaged across all layers on 35 concepts from the CustomConcept101 (Kumari et al. 2023) dataset. As in Fig. 4(c), FLLP exhibits 22% lower parameter drift, highlighting positive inter-concept interactions.

Conclusion

In this work, we introduced FLLP, a novel framework that leverages hyperbolic geometry and parent-child relational priors to alleviate catastrophic forgetting in CDMs. Unlike existing approaches that reduce inter-concept interactions, FLLP exploits structured inter-concept dependencies through a Lorentzian manifold, enabling positive transfer and robust continual adaptation. Across diverse benchmarks, our empirical results demonstrate that FLLP reliably enhances prior-knowledge retention, strengthens generalization, and scales effectively to new concepts.

References

- Anderson, J. W. 2006. *Hyperbolic geometry*. Springer Science & Business Media.
- Chen, X.; Huang, L.; Liu, Y.; Shen, Y.; Zhao, D.; and Zhao, H. 2024. AnyDoor: Zero-shot Object-level Image Cus-

- tomization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 6593–6602.
- Chen, Z.; and Liu, B. 2018. *Continual Learning and Catastrophic Forgetting*, 55–75. Cham: Springer International Publishing. ISBN 978-3-031-01581-6.
- Cheng, J.; Wu, F.; Tian, Y.; Wang, L.; and Tao, D. 2020. RiFeGAN: Rich Feature Generation for Text-to-Image Synthesis From Prior Knowledge. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10908–10917.
- Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255.
- Desai, K.; Nickel, M.; Rajpurohit, T.; Johnson, J.; and Vedantam, R. 2023. Hyperbolic image-text representations. In *Proceedings of the 40th International Conference on Machine Learning, ICML’23*. JMLR.org.
- Dong, J.; Liang, W.; Li, H.; Zhang, D.; Cao, M.; Ding, H.; Khan, S.; and Khan, F. S. 2024. How to continually adapt text-to-image diffusion models for flexible customization? In *Proceedings of the 38th International Conference on Neural Information Processing Systems, NIPS*. Red Hook, NY, USA: Curran Associates Inc. ISBN 9798331314385.
- Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; Uszkoreit, J.; and Houlsby, N. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ICLR*.
- Doyle, J.; and Rivest, R. L. 1976. Linear expected time of a simple union-find algorithm. *Information Processing Letters*, 5(5): 146–148.
- Fu, X.; Gao, Y.; Wei, Y.; Sun, Q.; Peng, H.; Li, J.; and Li, X. 2024. Hyperbolic geometric latent diffusion model for graph generation. In *Proceedings of the 41st International Conference on Machine Learning, ICML’24*. JMLR.org.
- Gal, R.; Alaluf, Y.; Atzmon, Y.; Patashnik, O.; Bermano, A. H.; Chechik, G.; and Cohen-Or, D. 2022. An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion. arXiv:2208.01618.
- Ganea, O.; Becigneul, G.; and Hofmann, T. 2018. Hyperbolic Entailment Cones for Learning Hierarchical Embeddings. In Dy, J.; and Krause, A., eds., *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, 1646–1655. PMLR.
- Gu, Y.; Wang, X.; Wu, J. Z.; Shi, Y.; Chen, Y.; Fan, Z.; Xiao, W.; Zhao, R.; Chang, S.; Wu, W.; Ge, Y.; Shan, Y.; and Shou, M. Z. 2023. Mix-of-Show: Decentralized Low-Rank Adaptation for Multi-Concept Customization of Diffusion Models. arXiv:2305.18292.
- Guo, Z.; and Jin, T. 2025. ConceptGuard: Continual Personalized Text-to-Image Generation with Forgetting and Confusion Mitigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2945–2954.
- Hu, E. J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; and Chen, W. 2022. LoRA: Low-Rank Adaptation of Large Language Models. In *International Conference on Learning Representations*.
- Kirkpatrick, J.; Pascanu, R.; Rabinowitz, N.; Veness, J.; Desjardins, G.; Rusu, A. A.; Milan, K.; Quan, J.; Ramalho, T.; Grabska-Barwinska, A.; Hassabis, D.; Clopath, C.; Kumaran, D.; and Hadsell, R. 2017. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13): 3521–3526.
- Kumari, N.; Zhang, B.; Zhang, R.; Shechtman, E.; and Zhu, J.-Y. 2023. Multi-Concept Customization of Text-to-Image Diffusion. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1931–1941.
- Law, M.; Liao, R.; Snell, J.; and Zemel, R. 2019. Lorentzian Distance Learning for Hyperbolic Representations. In Chaudhuri, K.; and Salakhutdinov, R., eds., *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, 3672–3681. PMLR.
- Le, M.; Roller, S.; Papaxanthos, L.; Kiela, D.; and Nickel, M. 2019. Inferring Concept Hierarchies from Text Corpora via Hyperbolic Embeddings. In Korhonen, A.; Traum, D.; and Márquez, L., eds., *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3231–3241. Florence, Italy: Association for Computational Linguistics.
- Li, Z.; and Hoiem, D. 2018. Learning without Forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(12): 2935–2947.
- Liu, X.; Zhang, R.; Zhao, J.; and Tan, M. 2023. HyperDiffusion: Generative modeling in hyperbolic space. *arXiv preprint arXiv:2302.05859*.
- Liu, Z.; Luo, P.; Wang, X.; and Tang, X. 2015. Deep Learning Face Attributes in the Wild. In *Proceedings of International Conference on Computer Vision (ICCV)*.
- Lokhande, V. S.; Chakraborty, R.; Ravi, S. N.; and Singh, V. 2022. Equivariance allows handling multiple nuisance variables when analyzing pooled neuroimaging datasets. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10432–10441.
- Lorentz, H. A.; Einstein, A.; Minkowski, H.; Weyl, H.; and Sommerfeld, A. 1923. *The principle of relativity: a collection of original memoirs on the special and general theory of relativity*. Dover.
- Park, T.; Liu, M.-Y.; Wang, T.-C.; and Zhu, J.-Y. 2019. Semantic Image Synthesis With Spatially-Adaptive Normalization. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2332–2341.
- Podell, D.; English, Z.; Lacey, K.; Blattmann, A.; Dockhorn, T.; Müller, J.; Penna, J.; and Rombach, R. 2024. SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis. In Kim, B.; Yue, Y.; Chaudhuri, S.; Fragkiadaki, K.; Khan, M.; and Sun, Y., eds., *International Conference on Representation Learning*, volume 2024, 1862–1874.

- Ramasinghe, S.; Shevchenko, V.; Avraham, G.; and Thalaiyasingam, A. 2024. Accept the Modality Gap: An Exploration in the Hyperbolic Space. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 27253–27262.
- Ramesh, A.; Dhariwal, P.; Nichol, A.; Chu, C.; and Chen, M. 2022. Hierarchical Text-Conditional Image Generation with CLIP Latents. *ArXiv*, abs/2204.06125.
- Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10684–10695.
- Ruiz, N.; Li, Y.; Jampani, V.; Pritch, Y.; Rubinstein, M.; and Aberman, K. 2023. DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 22500–22510.
- Sala, F.; De Sa, C.; Gu, A.; and Ré, C. 2018. Representation tradeoffs for hyperbolic embeddings. In *International conference on machine learning*, 4460–4469. PMLR.
- Smith, J. S.; Hsu, Y.; Zhang, L.; Hua, T.; Kira, Z.; Shen, Y.; and Jin, H. 2024. Continual Diffusion: Continual Customization of Text-to-Image Diffusion with C-LoRA. *Trans. Mach. Learn. Res.*
- Sun, G.; Liang, W.; Dong, J.; Li, J.; Ding, Z.; and Cong, Y. 2024. Create Your World: Lifelong Text-to-Image Diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(9): 6454–6470.
- Theves, S.; Neville, D.; Fernández, G.; and Doeller, C. 2021. Learning and Representation of Hierarchical Concepts in Hippocampus and Prefrontal Cortex. *The Journal of Neuroscience*, 41: JN–RM.
- Wang, K.; Yang, F.; Raducanu, B.; and van de Weijer, J. 2025. Multi-Class Textual-Inversion Secretly Yields a Semantic-Agnostic Classifier. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 4400–4409. Los Alamitos, CA, USA: IEEE Computer Society.
- Wang, X.; Chen, T.; Ge, Q.; Xia, H.; Bao, R.; Zheng, R.; Zhang, Q.; Gui, T.; and Huang, X. 2023. Orthogonal Subspace Learning for Language Model Continual Learning. In Bouamor, H.; Pino, J.; and Bali, K., eds., *Findings of the Association for Computational Linguistics: EMNLP*, 10658–10671. Singapore: Association for Computational Linguistics.
- Yu, Z.; Yang, Z.; and Zhang, J. 2024. DreamSteerer: enhancing source image conditioned editability using personalized diffusion models. In *Proceedings of the 38th International Conference on Neural Information Processing Systems, NIPS*. Red Hook, NY, USA: Curran Associates Inc. ISBN 9798331314385.
- Zhang, J.; Liu, L.; Silvén, O.; Pietikäinen, M.; and Hu, D. 2025. Few-Shot Class-Incremental Learning for Classification and Object Detection: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–20.
- Zhang, Y.; Huang, N.; Tang, F.; Huang, H.; Ma, C.; Dong, W.; and Xu, C. 2023. Inversion-based Style Transfer with Diffusion Models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10146–10156.
- Zhao, Y.; Luo, J.; Luo, X.; Huang, J.; Yuan, J.; Xiao, Z.; and Zhang, M. 2025. Attention bootstrapping for multimodal test-time adaptation. In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence, AAAI/IAAI/EAAI*. AAAI Press. ISBN 978-1-57735-897-8.
- Zhong, M.; Shen, Y.; Wang, S.; Lu, Y.; Jiao, Y.; Ouyang, S.; Yu, D.; Han, J.; and Chen, W. 2024. Multi-LoRA Composition for Image Generation. *Transactions on Machine Learning Research*.