

# Novelty Assessment Report

**Paper:** Directional Textual Inversion for Personalized Text-to-Image Generation

**PDF URL:** <https://openreview.net/pdf?id=6wA4qpyyU9>

**Venue:** ICLR 2026 Conference Submission

**Year:** 2026

**Report Generated:** 2026-01-07

## Abstract

Textual Inversion (TI) is an efficient approach to text-to-image personalization but often fails on complex prompts. We trace these failures to embedding norm inflation: learned tokens drift to out-of-distribution magnitudes, degrading prompt conditioning in pre-norm Transformers. Empirically, we show semantics are primarily encoded by direction in CLIP token space, while inflated norms harm contextualization; theoretically, we analyze how large magnitudes attenuate positional information and hinder residual updates in pre-norm blocks. We propose Directional Textual Inversion (DTI), which fixes the embedding magnitude to an in-distribution scale and optimizes only direction on the unit hypersphere via Riemannian SGD. We cast direction learning as MAP with a von Mises-Fisher prior, yielding a constant-direction prior gradient that is simple and efficient to incorporate. Across personalization tasks, DTI improves text fidelity over TI and TI-variants while maintaining subject similarity. Crucially, DTI's hyperspherical parameterization enables smooth, semantically coherent interpolation between learned concepts (slerp), a capability that is absent in standard TI. Our findings suggest that direction-only optimization is a robust and scalable path for prompt-faithful personalization.

### Disclaimer

This report is **AI-GENERATED** using Large Language Models and WisPaper (a scholar search engine). It analyzes academic papers' tasks and contributions against retrieved prior work. While this system identifies **POTENTIAL** overlaps and novel directions, **ITS COVERAGE IS NOT EXHAUSTIVE AND JUDGMENTS ARE APPROXIMATE**. These results are intended to assist human reviewers and **SHOULD NOT** be relied upon as a definitive verdict on novelty.

Note that some papers exist in multiple, slightly different versions (e.g., with different titles or URLs). The system may retrieve several versions of the same underlying work. The current automated pipeline does not reliably align or distinguish these cases, so human reviewers will need to disambiguate them manually.

If you have any questions, please contact: mingzhang23@m.fudan.edu.cn

## Core Task Landscape

This paper addresses: **personalized text-to-image generation with embedding optimization**

A total of **50 papers** were analyzed and organized into a taxonomy with **22 categories**.

### Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Embedding Learning and Optimization Strategies**
- **Encoder-Based Personalization**
- **Model Fine-Tuning Approaches**
- **Compositional and Controllable Generation**
- **Prompt Engineering and Refinement**
- **Specialized Generation Tasks and Modalities**

### Complete Taxonomy Tree

- personalized text-to-image generation with embedding optimization Survey Taxonomy
- Embedding Learning and Optimization Strategies
  - Textual Embedding Optimization
  - Single-Concept Textual Inversion (3 papers)
    - [1] An image is worth one word: Personalizing text-to-image generation using textual inversion (Gal, 2022) [View paper](#)
    - [4] Efficient Personalized Text-to-image Generation by Leveraging Textual Subspace (Du, 2024) [View paper](#)
    - [14] CatVersion: Concatenating Embeddings for Diffusion-Based Text-to-Image Personalization (Ruoyu Zhao, 2025) [View paper](#)
  - Constrained Embedding Optimization ★ (3 papers)
    - [0] Directional Textual Inversion for Personalized Text-to-Image Generation (Anon et al., 2026) [View paper](#)
    - [25] Core: Context-regularized text embedding learning for text-to-image personalization (Fei-ze Wu, 2025) [View paper](#)
    - [33] Cross initialization for personalized text-to-image generation (Yin Jian, 2023) [View paper](#)
  - Disentangled Embedding Learning (4 papers)
    - [2] Personalized Text-to-Image Generation with Attribute Disentanglement and Feature Embedding (Hou, 2025) [View paper](#)
    - [20] Decoupled textual embeddings for customized image generation (Bai Jinfeng, 2024) [View paper](#)
    - [24] Selectively informative description can reduce undesired embedding entanglements in text-to-image personalization (Jimyeong Kim, 2024) [View paper](#)
    - [38] DECOR:Decomposition and Projection of Text Embeddings for Text-to-Image Customization (Kim Jin Hwa, 2024) [View paper](#)
  - Latent Space and Multi-Modal Embedding Methods
  - Space-Time and Hierarchical Embeddings (1 papers)
    - [13] A Neural Space-Time Representation for Text-to-Image Personalization (Yuval Alaluf, 2023) [View paper](#)
  - Decomposed Visual-Textual Embeddings (2 papers)
    - [19] Highly Personalized Text Embedding for Image Manipulation by Stable Diffusion (Han Inhwa, 2023) [View paper](#)
    - [35] Beyond inserting: Learning identity embedding for semantic-fidelity personalized diffusion generation (Li Yang, 2024) [View paper](#)
  - Multi-Concept Embedding Composition (4 papers)
  - [12] Multi-Concept Customization of Text-to-Image Diffusion (Nupur Kumari, 2023) [View paper](#)
  - [26] -ECLIPSE: Multi-Concept Personalized Text-to-Image Diffusion Models by Leveraging CLIP Latent Space (M Patel, 2024) [View paper](#)
  - [39] Non-confusing generation of customized concepts in diffusion models (Lin Wang, 2024) [View paper](#)

- [49] FaR: Enhancing Multi-concept Text-to-Image Diffusion via Concept Fusion and Localized Refinement (Gia-Nghia Tran, 2025) [View paper](#)
- Encoder-Based Personalization
  - Multi-Modal Encoder Architectures (2 papers)
  - [7] Blip-diffusion: Pre-trained subject representation for controllable text-to-image generation and editing (Li, 2023) [View paper](#)
  - [8] ELITE: Encoding Visual Concepts into Textual Embeddings for Customized Text-to-Image Generation (Yu-xiang Wei, 2023) [View paper](#)
  - Image-to-Embedding Encoders (3 papers)
  - [11] EZIGen: Enhancing zero-shot subject-driven image generation with precise subject encoding and decoupled guidance (Zicheng Duan, 2024) [View paper](#)
  - [15] Taming encoder for zero fine-tuning image customization with text-to-image diffusion models (Jia, 2023) [View paper](#)
  - [47] DreamTuner: Single Image is Enough for Subject-Driven Generation (Hua Miao, 2023) [View paper](#)
  - Domain-Tuned Encoders (2 papers)
  - [21] Encoder-based Domain Tuning for Fast Personalization of Text-to-Image Models (Rinon Gal, 2023) [View paper](#)
  - [42] Domain-Agnostic Tuning-Encoder for Fast Personalization of Text-To-Image Models (Moab Arar, 2023) [View paper](#)
- Model Fine-Tuning Approaches
  - Full or Selective Parameter Fine-Tuning (2 papers)
  - [10] DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation (Ruiz, 2023) [View paper](#)
  - [45] DisenBooth: Identity-Preserving Disentangled Tuning for Subject-Driven Text-to-Image Generation (Chen Hong, 2023) [View paper](#)
  - Low-Rank Adaptation (4 papers)
  - [22] Navigating Text-To-Image Customization: From LyCORIS Fine-Tuning to Model Evaluation (Hsieh, 2023) [View paper](#)
  - [28] When stylegan meets stable diffusion: a w+ adapter for personalized image generation (X Li, 2024) [View paper](#)
  - [31] Styledrop: Text-to-image synthesis of any style (K Sohn, 2023) [View paper](#)
  - [44] When StyleGAN Meets Stable Diffusion: a Adapter for Personalized Image Generation (X Li, 2023) [View paper](#)
  - Preference and Feedback-Based Tuning (1 papers)
  - [32] Personalized Image Editing in Text-to-Image Diffusion Models via Collaborative Direct Preference Optimization (Connor Dunlop, 2025) [View paper](#)
- Compositional and Controllable Generation
  - Regional and Layout Control (2 papers)
  - [3] Region-aware text-to-image generation via hard binding and soft refinement (Chen Zhennan, 2024) [View paper](#)
  - [50] ALR-GAN: Adaptive Layout Refinement for Text-to-Image Synthesis (Tan Hong-chen, 2023) [View paper](#)
  - Attribute Binding and Alignment (3 papers)
  - [9] Mitigating Compositional Failures in Text-to-Image Models with Causal Text Embedding Refinement (Arman Zarei, 2025) [View paper](#)
  - [30] Attdreambooth: Towards text-aligned personalized text-to-image generation (Qing Li, 2024) [View paper](#)
  - [34] Fine-Grained Alignment and Noise Refinement for Compositional Text-to-Image Generation (Abdollahi Ali, 2025) [View paper](#)
- Prompt Engineering and Refinement
  - Interactive and Multi-Modal Prompting (1 papers)
  - [6] PromptCharm: Text-to-Image Generation through Multi-modal Prompting and Refinement (Wang Zhijie, 2024) [View paper](#)
  - Automated Prompt Optimization (4 papers)
  - [17] From reflection to perfection: Scaling inference-time optimization for text-to-image diffusion models via reflection tuning (Zhuo Le, 2025) [View paper](#)
  - [29] CAT-TPT: Class-Agnostic Text-based Test-time Prompt Tuning for Vision-Language Models (Youjia Zhang, 2025) [View paper](#)
  - [37] G-refine: A general refiner for text-to-image generation (C Li, 2024) [View paper](#)
  - [43] Divide, evaluate, and refine: Evaluating and improving text-to-image alignment with iterative vqa feedback (Singh, 2023) [View paper](#)
  - Adversarial and Safety-Oriented Prompt Methods (2 papers)
  - [36] Safe text-to-image generation: Simply sanitize the prompt embedding (Qiu, 2024) [View paper](#)
  - [46] RAT: Injecting Implicit Bias for Text-To-Image Prompt Refinement Models (Jiang Meng, 2024) [View paper](#)
- Specialized Generation Tasks and Modalities
  - Glyph and Text Rendering (1 papers)
  - [18] Refining text-to-image generation: Towards accurate training-free glyph-enhanced image generation (Sanyam Laxhanpal, 2025) [View paper](#)
  - High-Resolution and Quality Enhancement (1 papers)
  - [40] RefineNet: Enhancing Text-to-Image Conversion with High-Resolution and Detail Accuracy through Hierarchical Transformers and Progressive Refinement (Shi Fan, 2023) [View paper](#)
  - Alternative Generative Architectures (2 papers)
  - [5] Personalized Text-to-Image Generation with Auto-Regressive Models (SUN Kaiyue, 2025) [View paper](#)
  - [27] Realistic Image Generation from Text by Using BERT-Based Embedding (Sanghyuck Na, 2022) [View paper](#)
  - Cross-Domain and Harmonization Tasks (2 papers)
  - [16] ITI-Gen: Inclusive Text-to-Image Generation (Cheng Zhang, 2023) [View paper](#)
  - [23] Zero-shot image harmonization with generative model prior (Jianqi Chen, 2025) [View paper](#)
  - Dual-Embedding and End-to-End Training (2 papers)
  - [41] Cross initialization for face personalization of text-to-image models (Lianyu Pang, 2024) [View paper](#)
  - [48] End-to-end Training for Text-to-Image Synthesis using Dual-Text Embeddings (Mittal, 2025) [View paper](#)

## Narrative

Core task: personalized text-to-image generation with embedding optimization. The field organizes around several complementary strategies for adapting diffusion models to user-specific concepts. Embedding Learning and Optimization Strategies focus on refining textual or latent representations without modifying model weights, encompassing methods like textual inversion (Image Worth One Word[1]) and constrained optimization approaches that guide embeddings toward desired attributes. Encoder-Based Personalization leverages dedicated networks to map reference images into embedding spaces (e.g., Blip-diffusion[7], ELITE[8]), enabling rapid adaptation. Model Fine-Tuning Approaches adjust diffusion model parameters directly, exemplified by DreamBooth[10] and its variants,

trading inference speed for fidelity. Compositional and Controllable Generation addresses multi-concept scenarios and attribute disentanglement (Multi-Concept Customization[12], Attribute Disentanglement[2]), while Prompt Engineering and Refinement explores how textual guidance can be iteratively improved. Specialized Generation Tasks extend these techniques to domains like typography or video.

Within Embedding Learning, a particularly active line explores constrained optimization to balance identity preservation and editability. Directional Textual Inversion[0] sits in this cluster, emphasizing directional constraints during embedding refinement to maintain semantic coherence while personalizing concepts. Nearby works like Cross Initialization[33] propose alternative initialization strategies to accelerate convergence, and Core[25] investigates core embedding structures for robust personalization. In contrast, encoder-based methods (Taming Encoder[15], ECLIPSE[26]) prioritize zero-shot generalization by learning mappings from large datasets, sacrificing per-concept fine-grained control for speed. The tension between optimization depth and generalization breadth remains central: embedding optimization methods like Directional Textual Inversion[0] offer precise control over individual concepts but require iterative refinement, whereas encoder approaches achieve faster adaptation at the cost of potentially reduced fidelity for out-of-distribution subjects.

---

## Related Works in Same Category

---

The following **2 sibling papers** share the same taxonomy leaf node with the original paper:

### 1. Core: Context-regularized text embedding learning for text-to-image personalization

**Authors:** Fei-ze Wu, Yun Pang, Junyi Zhang, Lianyu Pang, Jian Yin, et al. (8 authors total) | **Year/Venue:** 2025 | **URL:** [View paper](#)

#### Abstract

Recent advances in text-to-image personalization have enabled high-quality and controllable image synthesis for user-provided concepts. However, existing methods still struggle to balance identity preservation with text alignment. Our approach is based on the fact that generating prompt-aligned images requires a precise semantic understanding of the prompt, which involves accurately processing the interactions between the new concept and its surrounding context tokens within the CLIP text encode...

#### Relationship Analysis

Both papers belong to the Constrained Embedding Optimization category, focusing on methods that impose constraints on embedding optimization to improve personalized text-to-image generation. They overlap in addressing the problem of embedding norm inflation and its negative impact on text fidelity, with both proposing regularization strategies to constrain the learned embeddings. The key difference is that the original paper (DTI) constrains embeddings to a fixed in-distribution magnitude and optimizes only direction on the unit hypersphere using Riemannian SGD with a von Mises-Fisher prior, while CoRe regularizes the context tokens surrounding the new concept by enforcing similarity constraints on their output embeddings and attention maps relative to a reference prompt containing a super-category token.

---

### 2. Cross initialization for personalized text-to-image generation

**Authors:** Yin Jian, Lianyu Pang, Xie, Haoran, Jian Yin, et al. (13 authors total) | **Year/Venue:** 2023 | **URL:** [View paper](#)

#### Abstract

Recently, there has been a surge in face personalization techniques, benefiting from the advanced capabilities of pretrained text-to-image diffusion models. Among these, a notable method is Textual Inversion, which generates personalized images by inverting given images into textual embeddings. However, methods based on Textual Inversion still struggle with balancing the trade-off between reconstruction quality and editability. In this study, we examine this issue through the lens of initializat...

#### Relationship Analysis

Both papers belong to the Constrained Embedding Optimization category, focusing on methods that impose constraints on embedding optimization for personalized text-to-image generation. The original paper (Directional Textual Inversion) constrains embeddings to a fixed in-distribution magnitude while optimizing only direction on the unit hypersphere using Riemannian SGD and a von Mises-Fisher prior, whereas the candidate paper (Cross Initialization) addresses embedding optimization through a novel initialization strategy that sets the initial embedding to the text encoder's output and applies L2 regularization to keep embeddings close to initialization. The key difference lies in their constraint mechanisms: the original enforces geometric constraints via hyperspherical optimization with directional priors, while the candidate enforces constraints through initialization-based regularization in Euclidean space.

---

## Contributions Analysis

---

**Overall novelty summary.** The paper proposes Directional Textual Inversion (DTI), which constrains embedding optimization to the unit hypersphere via Riemannian SGD and incorporates a von Mises-Fisher prior. It resides in the 'Constrained Embedding Optimization' leaf, which contains only three papers total. This leaf sits within the broader 'Textual Embedding Optimization' branch, indicating a moderately sparse research direction focused on geometric or semantic constraints during embedding learning. The small sibling set suggests this specific angle—directional constraints with hyperspherical parameterization—is relatively underexplored compared to unconstrained textual inversion methods.

The taxonomy tree shows that neighboring leaves include 'Single-Concept Textual Inversion' (three papers on unconstrained optimization) and 'Disentangled Embedding Learning' (four papers on identity-context separation). The 'Constrained Embedding Optimization' leaf explicitly excludes unconstrained methods and disentanglement-focused approaches, positioning DTI as a middle ground: it imposes geometric constraints without explicit disentanglement objectives. Nearby branches like 'Encoder-Based Personalization' and 'Model Fine-Tuning Approaches' represent alternative paradigms (feed-forward encoders vs. parameter updates), highlighting that DTI's iterative embedding refinement occupies a distinct methodological niche within the field.

Among twelve candidates examined, the MAP formulation with von Mises-Fisher prior (Contribution B) shows two refutable candidates out of ten examined, indicating some prior work on directional priors or hyperspherical embeddings. The DTI framework itself (Contribution A) examined two candidates with zero refutations, suggesting the specific combination of norm-fixing and Riemannian optimization may be novel. The theoretical analysis of norm inflation (Contribution C) was not tested against candidates. The limited search scope—twelve papers total—means these findings reflect top semantic matches rather than exhaustive coverage, and the refutable pairs likely represent overlapping methodological components rather than complete anticipation of the full DTI approach.

Given the sparse taxonomy leaf and the limited refutation rate across contributions, DTI appears to introduce a relatively fresh angle on constrained embedding optimization. However, the presence of two refutable candidates for the von Mises-Fisher prior suggests that directional priors are not entirely unprecedented. The analysis is bounded by the top-12 semantic search scope and does not cover the full landscape of hyperspherical learning or Riemannian optimization in adjacent fields.

---

This paper presents **3 main contributions**, each analyzed against relevant prior work:

### Contribution 1: Directional Textual Inversion (DTI) framework

**Description:** DTI is a novel personalization framework that decouples token embeddings into magnitude and direction components. It maintains embedding magnitude at in-distribution scale while optimizing only the directional component on the unit hypersphere using Riemannian SGD, improving text fidelity while preserving subject similarity.

This contribution was assessed against **2 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

### 1. DKD: Directional Knowledge Distillation for One-Step Text-to-Image Generation

URL: [View paper](#)

#### Brief Assessment

DKD[61] focuses on distilling multi-step diffusion models into one-step generators by modeling weight direction changes in U-Net/DiT architectures, not on personalizing text-to-image models through token embedding optimization on hyperspheres as in DTI.

---

### 2. Misalignment Attack on Text-to-Image Models via Text Embedding Optimization and Inversion

URL: [View paper](#)

#### Brief Assessment

Misalignment Attack[62] focuses on adversarial attacks via text embedding optimization to create misaligned text-image pairs, not on personalization frameworks that decouple embeddings into magnitude and direction for improved text fidelity.

---

## Contribution 2: MAP formulation with von Mises-Fisher prior for direction learning

**Description:** The authors formulate directional optimization as Maximum a Posteriori estimation with a von Mises-Fisher distribution as a directional prior. This yields a constant-direction prior gradient that regularizes embeddings towards semantically meaningful directions in hyperspherical latent space.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

### 1. Statistical modeling of directional data using a robust hierarchical von mises distribution model: perspectives for wind energy

URL: [View paper](#)

#### Brief Assessment

Robust Hierarchical vMF[57] focuses on statistical modeling of directional wind energy data using hierarchical von Mises distributions, not on embedding optimization or directional learning in neural networks for text-to-image generation.

---

### 2. Dynamic deep clustering of high-dimensional directional data via hyperspherical embeddings with Bayesian nonparametric mixtures

URL: [View paper](#)

#### Brief Assessment

Dynamic Deep Clustering[54] uses von Mises-Fisher mixtures for clustering directional data in VAE architectures, not for MAP-based embedding optimization in text-to-image personalization. The application domains and optimization objectives differ fundamentally.

---

### 3. Clustering on the Unit Hypersphere using von Mises-Fisher Distributions.

URL: [View paper](#)

#### Brief Assessment

Unit Hypersphere Clustering[55] focuses on clustering directional data using von Mises-Fisher distributions in an unsupervised setting, not on directional optimization for embedding learning with MAP formulation as a regularization technique for text-to-image personalization.

---

### 4. vMF-exp: von Mises-Fisher Exploration of Large Action Sets with Hyperspherical Embeddings

URL: [View paper](#)

#### Brief Assessment

vMF-exp[56] uses von Mises-Fisher distributions for exploration in reinforcement learning action spaces, not for directional optimization of text embeddings in generative models. The application domains and optimization objectives are fundamentally different.

---

### 5. Deep Adaptive Graph Clustering via von Mises-Fisher Distributions

URL: [View paper](#)

#### Brief Assessment

Adaptive Graph Clustering[51] uses von Mises-Fisher distributions for graph clustering tasks with node embeddings, not for directional optimization in text-to-image personalization or token embedding spaces.

---

### 6. Mises-Fisher similarity-based boosted additive angular margin loss for breast cancer classification

URL: [View paper](#)

#### Prior Art Analysis

Mises-Fisher Boosted Loss[60] demonstrates that von Mises-Fisher (VMF) distribution has been previously applied to directional optimization in deep learning contexts. The candidate paper explicitly formulates a loss function using VMF distribution to model directional features and control intra-class compactness through a concentration parameter  $\kappa$ , predating the original paper's MAP formulation. Both papers use VMF distribution to regularize embedding directions in hyperspherical space, with the candidate providing detailed mathematical formulations (equations 8-11) showing how VMF similarity replaces cosine similarity for directional learning.

#### Evidence

Evidence 1 - **Rationale:** Both papers use VMF distribution as a regularization mechanism for directional embeddings. The candidate explicitly describes how the concentration parameter  $\kappa$  controls compactness and serves as regularization, which is the same conceptual approach as the original's directional prior. - **Original:** this formulation incorporates a von mises-fisher (vmf) distribution as a directional prior, which effectively regularizes the embedding towards semantically meaningful directions in the hyperspherical latent space - **Candidate:** by controlling the parameter  $\kappa$ , the vmf similarity function exhibits distinctive properties compared to the cosine similarity function. specifically, a larger  $\kappa > 0$  induces a compact similarity function, which sensitively measures similarity around  $\theta = 0$ . from a classification perspective, this prop...

Evidence 2 - **Rationale:** Both papers emphasize computational efficiency while using VMF-based approaches. The candidate demonstrates that VMF similarity can be computationally simple and stable, supporting the claim that VMF-based directional learning was already established as an efficient approach. - **Original:** the resulting framework preserves the lightweight nature of ti while significantly improving its robustness, ensuring that personalization is both computationally efficient and semantically faithful - **Candidate:** the t-vmf similarity function ( 11) demonstrates a heavy tail that is better than the cosine tail, facilitating stable cnn training

even with varying  $\kappa$  values. this similarity function offers computational simplicity compared to vmf, which relies on an exponential function

---

## 7. vMFER: von Mises-Fisher Experience Resampling Based on Uncertainty of Gradient Directions for Policy Improvement of Actor-Critic Algorithms

URL: [View paper](#)

### Brief Assessment

vMFER[58] uses von Mises-Fisher distributions to measure uncertainty of gradient directions in reinforcement learning policy improvement, not for MAP-based directional optimization of token embeddings in text-to-image generation.

---

## 8. von Mises-Fisher Mixture Model-based Deep learning: Application to Face Verification

URL: [View paper](#)

### Prior Art Analysis

vMF Face Verification[53] demonstrates prior work that formulates directional optimization using von Mises-Fisher (vMF) distributions in a Maximum a Posteriori (MAP) framework. The candidate paper explicitly derives a vMF mixture loss from MAP estimation with vMF priors for directional feature learning, predating the original paper's claim. Both papers use vMF distributions as directional priors to regularize embeddings towards semantically meaningful directions, with the candidate providing the theoretical foundation (Eq. 3, 9) that the original paper builds upon.

### Evidence

Evidence 1 - **Rationale:** Both papers use the same vMF distribution parameterization (mean direction  $\mu$  and concentration  $\kappa$ ) on the unit sphere for directional priors. - **Original:** for the prior term,  $-\log p(v)$ , we use a von mises-fisher (vmf) distribution on the direction  $v$  (detailed justification in appendix c.2). the vmf distribution is a probability distribution on the  $(d - 1)$ -sphere, analogous to the gaussian distribution in euclidean space. it is parameterized by a mean  $\mu$  and concentration  $\kappa$ . - **Candidate:** for a  $d$ -dimensional random unit vector  $x = [x_1, \dots, x_d]^T \in \text{sd-1 Crd}$  (i.e.,  $\|x\|_2 = 1$ ), the density function of the vmf distribution is defined as [41]:  $vd(x|\mu, \kappa) = cd(\kappa) \exp(\kappa \mu^T x)$  where,  $\mu$  denotes the mean (with  $\|\mu\|_2 = 1$ ) and  $\kappa$  denotes the concentration parameter (with  $\kappa \geq 0$ ).

Evidence 2 - **Rationale:** Both papers derive gradients for the vMF prior that involve the constant-direction term  $\kappa \mu$ , showing the same mathematical approach to incorporating the prior. - **Original:** constant-direction prior gradient. a useful property is that the euclidean gradient of  $\log$ -prior is a constant:  $\nabla_v(-\kappa \mu^T v) = -\kappa \mu$ . practically, we just add this vector to the data gradient before projecting to the tangent space and retracting. - **Candidate:**  $\partial_l \theta \kappa = \kappa \sum_{j=1}^d (p_j - y_j) \mu_j x_j$ ;  $\partial_l \theta \mu_j = (p_j - y_j) \kappa x_j$

---

## 9. Non-convex Pose Graph Optimization in SLAM via Proximal Linearized Riemannian ADMM

URL: [View paper](#)

### Brief Assessment

Proximal Linearized ADMM[59] uses von Mises-Fisher distribution for modeling rotation constraints in SLAM pose graph optimization, not for directional learning in embedding optimization or text-to-image personalization.

---

## 10. Dino as a von mises-fisher mixture model

URL: [View paper](#)

### Brief Assessment

[Final Audit Failure] The model insisted on a refutation claim but failed to provide verifiable evidence after multiple retries. Marked as cannot\_refute for safety. Please manually verify the candidate text.

---

## Contribution 3: Theoretical and empirical analysis of embedding norm inflation

**Description:** The authors provide both theoretical analysis and empirical evidence showing that excessive embedding norms in standard Textual Inversion attenuate positional information and cause residual update stagnation in pre-norm Transformers, degrading text-prompt alignment.

This contribution was assessed against **0 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

## Appendix: Text Similarity Detection

Textual similarity detection checked 14 papers and found 1 similarity segment(s) across 1 paper(s).

The following **1 paper(s)** were detected to have high textual similarity with the original paper. These may represent different versions of the same work, duplicate submissions, or papers with substantial textual overlap. Readers are advised to verify these relationships independently.

### 1. Clustering on the Unit Hypersphere using von Mises-Fisher Distributions.

**Detected in:** contribution\_2

△ **Note:** This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

---

## References

- [0] Directional Textual Inversion for Personalized Text-to-Image Generation [View paper](#)
- [1] An image is worth one word: Personalizing text-to-image generation using textual inversion [View paper](#)
- [2] Personalized Text-to-Image Generation with Attribute Disentanglement and Feature Embedding [View paper](#)
- [3] Region-aware text-to-image generation via hard binding and soft refinement [View paper](#)
- [4] Efficient Personalized Text-to-image Generation by Leveraging Textual Subspace [View paper](#)
- [5] Personalized Text-to-Image Generation with Auto-Regressive Models [View paper](#)
- [6] PromptCharm: Text-to-Image Generation through Multi-modal Prompting and Refinement [View paper](#)
- [7] Blip-diffusion: Pre-trained subject representation for controllable text-to-image generation and editing [View paper](#)
- [8] ELITE: Encoding Visual Concepts into Textual Embeddings for Customized Text-to-Image Generation [View paper](#)
- [9] Mitigating Compositional Failures in Text-to-Image Models with Causal Text Embedding Refinement [View paper](#)
- [10] DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation [View paper](#)
- [11] EZIGen: Enhancing zero-shot subject-driven image generation with precise subject encoding and decoupled guidance [View paper](#)
- [12] Multi-Concept Customization of Text-to-Image Diffusion [View paper](#)
- [13] A Neural Space-Time Representation for Text-to-Image Personalization [View paper](#)
- [14] CatVersion: Concatenating Embeddings for Diffusion-Based Text-to-Image Personalization [View paper](#)

- [15] Taming encoder for zero fine-tuning image customization with text-to-image diffusion models [View paper](#)
- [16] ITI-Gen: Inclusive Text-to-Image Generation [View paper](#)
- [17] From reflection to perfection: Scaling inference-time optimization for text-to-image diffusion models via reflection tuning [View paper](#)
- [18] Refining text-to-image generation: Towards accurate training-free glyph-enhanced image generation [View paper](#)
- [19] Highly Personalized Text Embedding for Image Manipulation by Stable Diffusion [View paper](#)
- [20] Decoupled textual embeddings for customized image generation [View paper](#)
- [21] Encoder-based Domain Tuning for Fast Personalization of Text-to-Image Models [View paper](#)
- [22] Navigating Text-To-Image Customization: From LyCORIS Fine-Tuning to Model Evaluation [View paper](#)
- [23] Zero-shot image harmonization with generative model prior [View paper](#)
- [24] Selectively informative description can reduce undesired embedding entanglements in text-to-image personalization [View paper](#)
- [25] Core: Context-regularized text embedding learning for text-to-image personalization [View paper](#)
- [26] -ECLIPSE: Multi-Concept Personalized Text-to-Image Diffusion Models by Leveraging CLIP Latent Space [View paper](#)
- [27] Realistic Image Generation from Text by Using BERT-Based Embedding [View paper](#)
- [28] When stylegan meets stable diffusion: a w+ adapter for personalized image generation [View paper](#)
- [29] CAT-TPT: Class-Agnostic Text-based Test-time Prompt Tuning for Vision-Language Models [View paper](#)
- [30] Atndreambooth: Towards text-aligned personalized text-to-image generation [View paper](#)
- [31] Styledrop: Text-to-image synthesis of any style [View paper](#)
- [32] Personalized Image Editing in Text-to-Image Diffusion Models via Collaborative Direct Preference Optimization [View paper](#)
- [33] Cross initialization for personalized text-to-image generation [View paper](#)
- [34] Fine-Grained Alignment and Noise Refinement for Compositional Text-to-Image Generation [View paper](#)
- [35] Beyond inserting: Learning identity embedding for semantic-fidelity personalized diffusion generation [View paper](#)
- [36] Safe text-to-image generation: Simply sanitize the prompt embedding [View paper](#)
- [37] G-refine: A general refiner for text-to-image generation [View paper](#)
- [38] DECOR:Decomposition and Projection of Text Embeddings for Text-to-Image Customization [View paper](#)
- [39] Non-confusing generation of customized concepts in diffusion models [View paper](#)
- [40] RefineNet: Enhancing Text-to-Image Conversion with High-Resolution and Detail Accuracy through Hierarchical Transformers and Progressive Refinement [View paper](#)
- [41] Cross initialization for face personalization of text-to-image models [View paper](#)
- [42] Domain-Agnostic Tuning-Encoder for Fast Personalization of Text-To-Image Models [View paper](#)
- [43] Divide, evaluate, and refine: Evaluating and improving text-to-image alignment with iterative vqa feedback [View paper](#)
- [44] When StyleGAN Meets Stable Diffusion: a Adapter for Personalized Image Generation [View paper](#)
- [45] DisenBooth: Identity-Preserving Disentangled Tuning for Subject-Driven Text-to-Image Generation [View paper](#)
- [46] RAT: Injecting Implicit Bias for Text-To-Image Prompt Refinement Models [View paper](#)
- [47] DreamTuner: Single Image is Enough for Subject-Driven Generation [View paper](#)
- [48] End-to-end Training for Text-to-Image Synthesis using Dual-Text Embeddings [View paper](#)
- [49] FaR: Enhancing Multi-concept Text-to-Image Diffusion via Concept Fusion and Localized Refinement [View paper](#)
- [50] ALR-GAN: Adaptive Layout Refinement for Text-to-Image Synthesis [View paper](#)
- [51] Deep Adaptive Graph Clustering via von Mises-Fisher Distributions [View paper](#)
- [52] Dino as a von mises-fisher mixture model [View paper](#)
- [53] von Mises-Fisher Mixture Model-based Deep learning: Application to Face Verification [View paper](#)
- [54] Dynamic deep clustering of high-dimensional directional data via hyperspherical embeddings with Bayesian nonparametric mixtures [View paper](#)
- [55] Clustering on the Unit Hypersphere using von Mises-Fisher Distributions. [View paper](#)
- [56] vMF-exp: von Mises-Fisher Exploration of Large Action Sets with Hyperspherical Embeddings [View paper](#)
- [57] Statistical modeling of directional data using a robust hierarchical von mises distribution model: perspectives for wind energy [View paper](#)
- [58] vMFER: von Mises-Fisher Experience Resampling Based on Uncertainty of Gradient Directions for Policy Improvement of Actor-Critic Algorithms [View paper](#)
- [59] Non-convex Pose Graph Optimization in SLAM via Proximal Linearized Riemannian ADMM [View paper](#)
- [60] Mises-Fisher similarity-based boosted additive angular margin loss for breast cancer classification [View paper](#)
- [61] DKD: Directional Knowledge Distillation for One-Step Text-to-Image Generation [View paper](#)
- [62] Misalignment Attack on Text-to-Image Models via Text Embedding Optimization and Inversion [View paper](#)