Novelty Assessment Report

Paper: WithAnyone: Toward Controllable and ID Consistent Image Generation
PDF URL: https://openreview.net/pdf?id=xFo13SaHQm
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2026-01-05

Abstract

Identity-consistent (ID-consistent) generation has become an important focus in text-to-image research, with recent models achieving notable success in producing images aligned with a reference identity. Yet, the scarcity of large-scale paired datasets—containing multiple images of the same individual—forces most approaches to adopt reconstruction-based training. This reliance often leads to a failure mode we term copy-paste, where the model directly replicates the reference face rather than preserving identity across natural variations in pose, expression, or lighting. Such over-similarity undermines controllability and limits the expressive power of generation. To address these limitations, we (1) construct a large-scale paired dataset, MultiID-2M, tailored for multi-person scenarios, providing diverse references for each identity; (2) introduce a benchmark that quantifies both copy-paste artifacts and the trade-off between identity fidelity and variation; and (3) propose a novel training paradigm with a contrastive identity loss that leverages paired data to balance fidelity with diversity. These contributions culminate in WithAnyone, a diffusion-based model that effectively mitigates copy-paste while preserving high identity similarity. Extensive experiments—both qualitative and quantitative—demonstrate that WithAnyone substantially reduces copy-paste artifacts, improves controllability over pose and expression, and maintains strong perceptual quality. User studies further validate that our method achieves high identity fidelity while enabling expressive, controllable generation.

Core Task Landscape

This paper addresses: **Identity-Consistent Image Generation with Controllable Variations**

A total of **50 papers** were analyzed and organized into a taxonomy with **19 categories**.

Taxonomy Overview

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

- **Identity-Preserving Text-to-Image Generation**
- **Identity-Consistent Video Generation**
- **Controllable Human Image Synthesis**
- **Cross-Modal Identity Transfer**
- **Synthetic Identity Generation for Recognition**
- **Disentangled Representation Learning**
- **Spatial and Compositional Control**
- **Specialized Identity-Consistent Applications**
- **Identity-Controllability Trade-off Optimization**

Complete Taxonomy Tree

- Identity-Consistent Image Generation with Controllable Variations Survey Taxonomy
- Identity-Preserving Text-to-Image Generation
  - Tuning-Free Identity Preservation (4 papers)
  - [6] Blip-diffusion: Pre-trained subject representation for controllable text-to-image generation and editing (Li, 2023) View paper
  - [18] Training-Free Consistent Text-to-Image Generation (Yoad Tewel, 2024) View paper
  - [20] MagicNaming: Consistent Identity Generation by Finding a "Name Space" in T2I Diffusion Models (Huang, 2024) View paper
  - [39] Taming Encoder for Zero Fine-tuning Image Customization with Text-to-Image Diffusion Models (Jia, 2023) View paper
  - Fine-Tuning-Based Identity Adaptation (3 papers)
  - [3] ID-Booth: Identity-consistent Face Generation with Diffusion Models (TomaÅ¡eviÄ Darian, 2025) View paper
  - [38] PortraitBooth: A Versatile Portrait Model for Fast Identity-Preserved Personalization (Xu Peng, 2023) View paper
  - [44] FaceChain-FACT: Face Adapter with Decoupled Training for Identity-preserved Personalization (Yu Cheng, 2024) View paper
  - Multi-Subject and Story Generation (5 papers)
  - [9] LayerCraft: Enhancing Text-to-Image Generation with CoT Reasoning and Layered Object Integration (ZHANG Yuyao, 2025) View paper
  - [10] One-Prompt-One-Story: Free-Lunch Consistent Text-to-Image Generation Using a Single Prompt (Liu, 2025) View paper
  - [35] Disco: Reinforcement with diversity constraints for multi-human generation (Borse, 2025) View paper
  - [42] AutoStory: Generating Diverse Storytelling Images with Minimal Human Efforts (Wen Wang, 2024) View paper
  - [47] StoryMaker: Towards Holistic Consistent Characters in Text-to-image Generation (Zhou Zheng-guang, 2024) View paper
- Identity-Consistent Video Generation
  - Text-to-Video with Identity Preservation (3 papers)
  - [1] Identity-preserving text-to-video generation by frequency decomposition (Shenghai Yuan, 2025) View paper
  - [16] Motioncharacter: Identity-preserving and motion controllable human video generation (Qiu Di, 2024) View paper
  - [29] Concat-ID: Towards Universal Identity-Preserving Video Synthesis (Zhong Yong, 2025) View paper
  - Image-to-Video Animation (3 papers)

- [4] Motion-I2V: Consistent and Controllable Image-to-Video Generation with Explicit Motion Modeling (Xiaoyu Shi, 2024) View paper
- [5] Animate Anyone: Consistent and Controllable Image-to-Video Synthesis for Character Animation (H. Li, 2024) View paper
- [45] Frame In-N-Out: Unbounded Controllable Image-to-Video Generation (Wang Boyang, 2025) View paper

- Controllable Human Image Synthesis
  - Pose and Expression Control (5 papers)
  - [8] ID-Consistent, Precise Expression Generation with Blendshape-Guided Diffusion (Papantoniou, 2025) View paper
  - [13] Consistent Human Image and Video Generation with Spatially Conditioned Diffusion (Cao, 2024) View paper
  - [28] CP-EB: Talking Face Generation with Controllable Pose and Eye Blinking Embedding (Jianzong Wang, 2023) View paper
  - [48] Controlling Avatar Diffusion with Learnable Gaussian Embedding (Gao Xuan, 2025) View paper
  - [50] MagicPose: Realistic Human Poses and Facial Expressions Retargeting with Identity-aware Diffusion (Chang Di, 2023) View paper
  - Garment and Appearance Control (3 papers)
  - [7] Controllable Human Image Generation with Personalized Multi-Garments (Yisol Choi, 2025) View paper
  - [14] ComposeMe: Attribute-Specific Image Prompts for Controllable Human Image Generation (Guocheng Qian, 2025) View paper
  - [22] From parts to whole: A unified reference framework for controllable human image generation (Huang Zehuan, 2024) View paper
  - Attribute-Specific Modular Control (2 papers)
  - [27] FlexIP: Dynamic Control of Preservation and Personality for Customized Image Generation (Huang Lin-yan, 2025) View paper
  - [31] Omni-id: Holistic identity representation designed for generative tasks (Guocheng Qian, 2025) View paper
- Cross-Modal Identity Transfer
  - Thermal-to-Visible Face Translation (2 papers)
  - [2] Identity-aware infrared person image generation and re-identification via controllable diffusion model (Xizhuo Yu, 2025) View paper
  - [26] DiffTV: Identity-preserved thermal-to-visible face translation via feature alignment and dual-stage conditions (Jingyu Lin, 2024) View paper
  - Multi-Modal Person Re-Identification (1 papers)
  - [25] Unified Conditional Image Generation for Visible-Infrared Person Re-Identification (Honghu Pan, 2024) View paper
- Synthetic Identity Generation for Recognition
  - Diverse Synthetic Face Generation (4 papers)
  - [19] GANDiffFace: Controllable Generation of Synthetic Datasets for Face Recognition with Realistic Variations (Pietro Melzi, 2023) View paper
  - [23] Idiff-face: Synthetic-based face recognition through fizzy identity-conditioned diffusion model (F Boutros, 2023) View paper
  - [32] ID: Identity-Preserving-yet-Diversified Diffusion Models for Synthetic Face Recognition (S Li, 2024) View paper
  - [40] FLUXSynID: A Framework for Identity-Controlled Synthetic Face Generation with Document and Live Images (Spreeuwers, 2025) View paper
  - Fingerprint Synthesis with Identity Control (1 papers)
  - [33] Universal Fingerprint Generation: Controllable Diffusion Model With Multimodal Conditions (Grosz, 2024) View paper
- Disentangled Representation Learning
  - 3D-Guided Disentanglement (2 papers)
  - [11] Disentangled and Controllable Face Image Generation via 3D Imitative-Contrastive Learning (Yu Deng, 2020) View paper
  - [36] Towards open-set identity preserving face synthesis (Jianmin Bao, 2018) View paper
  - Latent Space Manipulation (3 papers)
  - [15] Tedigan: Text-guided diverse face image generation and manipulation (Xia, 2021) View paper
  - [43] Linearly controllable gan: Unsupervised feature categorization and decomposition for image generation and manipulation (Sehyung Lee, 2024) View paper
  - [49] Face Identity-Aware Disentanglement in StyleGAN (Adrian SuwaÅ□a, 2024) View paper
  - Person Image Disentanglement (1 papers)
  - [24] Disentangled representation learning for controllable person image generation (Wenju Xu, 2024) View paper
- Spatial and Compositional Control
  - Multi-Subject Spatial Positioning (1 papers)
  - [37] PositionIC: Unified Position and Identity Consistency for Image Customization (Hu Junjie, 2025) View paper
  - Consistent Editing and Synthesis (3 papers)
  - [12] MasaCtrl: Tuning-Free Mutual Self-Attention Control for Consistent Image Synthesis and Editing (Mingdeng Cao, 2023) View paper
  - [17] CharaConsist: Fine-Grained Consistent Character Generation (Wang Mengyu, 2025) View paper
  - [30] Editable Image Elements for Controllable Synthesis (Mu Jiteng, 2024) View paper
- Specialized Identity-Consistent Applications (3 papers)
  - [21] Identityâ□□consistent transfer learning of portraits for digital apparel sample display (Luyuan Wang, 2024) View paper
  - [34] SINGAPO: Single Image Controlled Generation of Articulated Parts in Objects (Liu Jiayi, 2024) View paper
  - [41] 3D Cartoon Face Generation with Controllable Expressions from a Single GAN Image (Hao Wang, 2022) View paper
- Identity-Controllability Trade-off Optimization ★ (2 papers)
  - [0] WithAnyone: Toward Controllable and ID Consistent Image Generation (Anon et al., 2026) View paper
  - [46] WithAnyone: Towards Controllable and ID Consistent Image Generation (Xu Hengyuan, 2025) View paper

Narrative

Core task: Identity-consistent image generation with controllable variations. The field addresses the challenge of synthesizing images that preserve a subject's identity while allowing flexible control over pose, expression, scene composition, and other attributes. The taxonomy reveals a rich landscape organized around several complementary directions. Identity-Preserving Text-to-Image Generation focuses on injecting reference identities into diffusion models via text prompts, often using adapter modules or embedding techniques (e.g., Blip-diffusion[6], PortraitBooth[38]). Identity-Consistent Video Generation extends these ideas to temporal sequences, ensuring frame-to-frame coherence (e.g., Animate Anyone[5], Motion-I2V[4]). Controllable Human Image Synthesis emphasizes pose and garment transfer, while Cross-Modal Identity Transfer tackles domain shifts such as photo-to-sketch or infrared modalities (Infrared Person Generation[2]). Disentangled Representation Learning seeks to separate identity from other factors in latent space (Tedigan[15]), and

Spatial and Compositional Control provides fine-grained layout or multi-subject orchestration (MasaCtrl[12], LayerCraft[9]). Specialized applications target niche domains like cartoon faces or digital apparel, and Synthetic Identity Generation supports recognition tasks by creating diverse training data.

A central tension across these branches is the identity-controllability trade-off: stronger identity preservation can limit flexibility in pose, lighting, or scene composition, while aggressive control may dilute recognizable features. Recent works explore adaptive weighting schemes, frequency-domain decomposition (Frequency Decomposition[1]), and multi-stage pipelines to balance these competing goals. WithAnyone[0] sits squarely in the Identity-Controllability Trade-off Optimization branch, addressing how to maintain fidelity across varied conditions without sacrificing creative freedom. It shares thematic concerns with ID-Booth[3], which also targets robust identity encoding under diverse prompts, and contrasts with more application-specific methods like Multi-Garments[7] or Blendshape-Guided[8] that prioritize domain constraints over general trade-off tuning. This positioning highlights an ongoing effort to develop principled mechanisms—whether through loss formulations, architectural choices, or training strategies—that let practitioners dial identity strength and controllability according to downstream needs.

Related Works in Same Category

The following **1 sibling papers** share the same taxonomy leaf node with the original paper:

1. WithAnyone: Towards Controllable and ID Consistent Image Generation

Authors: Xu Hengyuan, Cheng Wei, Xing Peng, Wu Shuhan, Wang Rui, et al. (12 authors total) | **Year/Venue**: 2025 • arXiv.org | **URL**: View paper

Abstract

Identity-consistent generation has become an important focus in text-to-image research, with recent models achieving notable success in producing images aligned with a reference identity. Yet, the scarcity of large-scale paired datasets containing multiple images of the same individual forces most approaches to adopt reconstruction-based training. This reliance often leads to a failure mode we term copy-paste, where the model directly replicates the reference face rather than preserving identity...

⚠ Similarity Notice

This paper is highly similar to the original paper; it may be a variant or near-duplicate. Please manually verify.

Contributions Analysis

Overall novelty summary. The paper introduces WithAnyone, a diffusion model targeting identity-consistent generation with controllable variations, and contributes the MultiID-2M dataset plus a contrastive identity loss. It resides in the Identity-Controllability Trade-off Optimization leaf, which contains only two papers in the entire 50-paper taxonomy. This sparse positioning suggests the explicit framing of the identity-fidelity versus variation balance as a core optimization problem is relatively underexplored. Most prior work either prioritizes strong identity preservation (e.g., tuning-free or fine-tuning-based methods) or emphasizes attribute control (pose, garment) without systematically addressing the trade-off itself.

The taxonomy reveals neighboring directions: Identity-Preserving Text-to-Image Generation (tuning-free and fine-tuning branches with seven papers) focuses on embedding identities into diffusion models, while Controllable Human Image Synthesis (ten papers across pose, garment, and attribute control) emphasizes manipulation without explicit trade-off optimization. Disentangled Representation Learning (six papers) separates identity from attributes in latent space but does not directly tackle the copy-paste failure mode described here. WithAnyone's contrastive loss and paired-data paradigm diverge from these by explicitly penalizing over-similarity, a mechanism not prominent in sibling or adjacent categories.

Among 30 candidates examined, none clearly refute the three contributions. The MultiID-2M dataset (10 candidates, 0 refutable) appears novel in its multi-person paired structure tailored for identity diversity. The ID-contrastive training approach (10 candidates, 0 refutable) leverages paired data in a way not documented among the examined papers, though the limited search scope means exhaustive dataset or loss-function surveys were not conducted. The WithAnyone model (10 candidates, 0 refutable) integrates these elements into a unified framework, with no examined work presenting an equivalent combination of dataset, loss, and benchmark for the copy-paste problem.

Based on top-30 semantic matches and the sparse taxonomy leaf, the work appears to occupy a distinct niche. The explicit focus on the identity-controllability trade-off, paired-data training, and copy-paste quantification are not prominently addressed in the examined literature. However, the limited search scope and the small number of papers in the target leaf mean a broader survey could reveal additional relevant methods or datasets. The analysis covers the immediate semantic neighborhood but does not claim exhaustive coverage of all identity-consistent generation research.

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

Contribution 1: MultiID-2M dataset for multi-person ID-consistent generation

Description: The authors build a large-scale paired dataset called MultiID-2M specifically designed for multi-person scenarios. This dataset addresses the scarcity of paired data containing multiple images of the same individual, enabling better training for identity-consistent generation.

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. CustomVideo: Customizing Text-to-Video Generation with Multiple Subjects

URL: View paper

Brief Assessment

CustomVideo[55] focuses on multi-subject text-to-video generation with a video dataset, while the original paper presents MultiID-2M for multi-person image generation with paired identity references. These are fundamentally different modalities (video vs. image) and dataset construction approaches.

---

2. PortraitBooth: A Versatile Portrait Model for Fast Identity-Preserved Personalization

URL: View paper

Brief Assessment

PortraitBooth[38] focuses on single-subject portrait personalization using face recognition models and does not present a large-scale paired dataset for multi-person scenarios. The candidate's training approach uses standard face datasets without the paired multi-identity structure described in the original paper.

---

3. ContextGen: Contextual Layout Anchoring for Identity-Consistent Multi-Instance Generation

URL: View paper

Brief Assessment

ContextGen[51] focuses on multi-instance image generation with layout control and introduces the IMIG-100K dataset, which addresses layout and identity annotations for multiple objects. This differs from MultiID-2M, which specifically targets multi-person scenarios with paired celebrity reference images for identity-consistent generation across diverse expressions and poses.

---

4. Trackformer: Multi-object tracking with transformers

URL: View paper

Brief Assessment

Trackformer[53] addresses multi-object tracking in videos using transformers, not identity-consistent image generation or dataset construction for such tasks. The candidate focuses on tracking objects across video frames rather than generating images with consistent identities.

---

5. Id-patch: Robust id association for group photo personalization

URL: View paper

Brief Assessment

Id-patch[57] focuses on a different technical approach (ID patches and spatial control for group photo generation) rather than dataset construction. While it mentions using training data with multi-person images, it does not claim to build or release a large-scale paired dataset specifically designed for multi-person identity-consistent generation like MultiID-2M.

---

6. DanceTogether! Identity-Preserving Multi-Person Interactive Video Generation

URL: View paper

Brief Assessment

DanceTogether[52] focuses on controllable video generation with pose-mask streams for multi-actor interaction, not on building large-scale paired datasets for identity-consistent image generation. Their datasets (PairFS-4K, HumanRob-300) serve video generation tasks with pose control, which is a different technical domain from the static multi-person image generation that MultiID-2M addresses.

---

7. MultiHuman-Testbench: Benchmarking Image Generation for Multiple Humans

URL: View paper

Brief Assessment

MultiHuman-Testbench[54] is a benchmark for evaluation, not a training dataset. It contains 1,800 test samples with 5,550 reference faces for benchmarking purposes, whereas MultiID-2M is a large-scale training dataset with 500k paired images and 1.5m unpaired images designed for model training.

---

8. Concat-ID: Towards Universal Identity-Preserving Video Synthesis

URL: View paper

Brief Assessment

Concat-ID[29] focuses on identity-preserving video generation using variational autoencoders and 3d self-attention mechanisms, not on constructing large-scale paired datasets for multi-person image generation. The candidate paper does not discuss dataset construction methodologies or multi-person paired reference data collection.

---

9. Diffusion Self-Distillation for Zero-Shot Customized Image Generation

URL: View paper

Brief Assessment

Diffusion Self-Distillation[56] focuses on self-distillation techniques for identity-preserving generation using synthetic paired data from pre-trained models, not on constructing large-scale multi-person datasets with real celebrity references and identity labels.

---

10. WithAnyone: Towards Controllable and ID Consistent Image Generation

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 2: ID-contrastive training approach

Description: The authors introduce an ID-contrastive training method that helps the model preserve identity across natural variations in pose, expression, and lighting, while avoiding the copy-paste failure mode where models directly replicate reference faces.

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. Audio-driven emotion-aware 3d talking face generation from single image

URL: View paper

Brief Assessment

Audio-driven Emotion[61] focuses on audio-driven 3D talking face generation with emotion control, not identity preservation across pose/expression variations in text-to-image generation. The technical domains and objectives are fundamentally different.

---

2. Identity-aware convolutional neural network for facial expression recognition

URL: View paper

Brief Assessment

Identity-aware CNN[65] focuses on facial expression recognition with identity-invariant features, while the original paper addresses identity-consistent image generation. The candidate's contrastive loss aims to make expression features invariant to identity, whereas the original's ID-contrastive loss preserves identity across pose/expression variations—fundamentally opposite objectives.

---

3. Face Swapping via Reverse Contrastive Learning and Explicit Identity-Attribute Disentanglement

URL: View paper

Brief Assessment

Reverse Contrastive[64] focuses on face swapping tasks with reverse contrastive learning for identity-attribute disentanglement, while the original paper addresses multi-identity image generation with ID-contrastive loss using extended negative pools from paired datasets. The technical approaches and application domains differ substantially.

---

4. Cross-Age Contrastive Learning for Age-Invariant Face Recognition

URL: View paper

Brief Assessment

Cross-Age Contrastive[60] focuses on age-invariant face recognition using contrastive learning to handle age variations, not on general identity preservation across pose/expression variations in text-to-image generation.

---

5. Sample-cohesive pose-aware contrastive facial representation learning

URL: View paper

Brief Assessment

Sample-cohesive[59] focuses on pose-disentangled contrastive learning for facial representation, not identity preservation across variations. The candidate addresses pose-aware vs. non-pose features in facial analysis, while the original paper targets identity consistency in text-to-image generation with reference faces.

---

6. Real-time, high-fidelity face identity swapping with a vision foundation model

URL: View paper

Brief Assessment

Face Identity Swapping[62] uses contrastive losses with CLIP features for face swapping tasks, while the original paper applies ID-contrastive training with ArcFace embeddings for text-to-image generation. The technical contexts and objectives differ substantially.

---

7. Pose-disentangled Contrastive Learning for Self-supervised Facial Representation

URL: View paper

Brief Assessment

Pose-disentangled Contrastive[58] focuses on disentangling pose-related features from facial features for self-supervised learning, not on identity preservation across variations. The candidate addresses pose-invariance in facial representation, while the original paper tackles identity consistency in generative models.

---

8. Contrastive viewpoint-aware shape learning for long-term person re-identification

URL: View paper

Brief Assessment

Viewpoint-aware Shape[63] focuses on contrastive learning for viewpoint-aware shape features in person re-identification, not on preventing copy-paste artifacts in face generation across pose/expression variations as in the original paper's ID-contrastive training.

---

9. Disentangled and Controllable Face Image Generation via 3D Imitative-Contrastive Learning

URL: View paper

Brief Assessment

Imitative-Contrastive[11] focuses on disentangling face generation factors (identity, expression, pose, illumination) using 3D morphable models and contrastive learning between generated image pairs. The ORIGINAL paper's ID-contrastive training preserves identity across natural variations in multi-person scenarios using paired reference datasets, which is a fundamentally different application domain and technical approach.

---

10. Supervised contrastive learning with identity-label embeddings for facial action unit recognition

URL: View paper

Brief Assessment

Identity-Label Embeddings[66] focuses on facial action unit (AU) recognition using identity-label embeddings to handle individual facial variations, not on identity-consistent image generation across pose/expression variations as in the original paper's ID-contrastive training for diffusion models.

Contribution 3: WithAnyone model for controllable ID-consistent generation

Description: The authors develop WithAnyone, a model that generates high-quality images with controllable attributes while maintaining identity consistency. The model addresses limitations of reconstruction-based training that lead to over-similarity and reduced controllability.

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. SmartAvatar: Text- and Image-Guided Human Avatar Generation with VLM AI Agents

URL: View paper

Brief Assessment

SmartAvatar[69] focuses on 3D avatar generation from photos/text using VLM agents and parametric human models, not 2D image generation with identity consistency. The technical approaches are fundamentally different - WithAnyone addresses 2D diffusion-based image synthesis with identity preservation, while SmartAvatar[69] tackles 3D rigged avatar creation through procedural generation.

---

2. MasaCtrl: Tuning-Free Mutual Self-Attention Control for Consistent Image Synthesis and Editing

URL: View paper

Brief Assessment

MasaCtrl[12] focuses on consistent image generation through mutual self-attention mechanisms in diffusion models, addressing view/pose consistency and non-rigid editing. It does not address identity-consistent generation with controllable attributes or the copy-paste artifact problem that WithAnyone tackles through contrastive training and paired datasets.

---

3. Uniportrait: A unified framework for identity-preserving single-and multi-human image personalization

URL: View paper

Brief Assessment

Uniportrait[68] focuses on a unified framework for single- and multi-human image personalization with emphasis on layout generation and routing mechanisms, while WithAnyone specifically addresses copy-paste artifacts through paired training and contrastive learning on the MultiID-2M dataset. The technical approaches and problem formulations differ substantially.

---

4. Identity-aware infrared person image generation and re-identification via controllable diffusion model

URL: View paper

Brief Assessment

Infrared Person Generation[2] focuses on cross-modal infrared-to-visible person image generation and re-identification, which is a fundamentally different domain from WithAnyone's controllable identity-consistent generation in natural RGB images. The technical challenges and methods are domain-specific.

---

5. Motioncharacter: Identity-preserving and motion controllable human video generation

URL: View paper

Brief Assessment

Motioncharacter[16] focuses on identity-preserving human video generation with motion control, not controllable image generation with attribute control as in WithAnyone.

---

6. Animate Anyone: Consistent and Controllable Image-to-Video Synthesis for Character Animation

URL: View paper

Brief Assessment

Animate Anyone[5] focuses on character animation from still images using pose-driven video generation, not on controllable identity-consistent image generation with attribute control. The candidate addresses temporal consistency in video frames for character animation, while the original paper addresses copy-paste artifacts and controllability in multi-identity image generation scenarios.

---

7. ID-Booth: Identity-consistent Face Generation with Diffusion Models

URL: View paper

Brief Assessment

ID-Booth[3] focuses on fine-tuning diffusion models for identity-consistent face generation using a triplet identity loss, primarily for privacy-preserving dataset augmentation. The candidate does not address the copy-paste artifact problem or the controllability-fidelity trade-off that WithAnyone specifically targets through paired-data training and contrastive learning with extended negatives.

---

8. FLUXSynID: A Framework for Identity-Controlled Synthetic Face Generation with Document and Live Images

URL: View paper

Brief Assessment

FLUXSynID[40] focuses on synthetic face dataset generation with identity attribute control for biometric research applications, while the original paper addresses controllable identity-consistent image generation in text-to-image models with emphasis on mitigating copy-paste artifacts through contrastive training.

---

9. Ominicontrol: Minimal and universal control for diffusion transformer

URL: View paper

Brief Assessment

Ominicontrol[67] focuses on general image conditioning for diffusion transformers across spatially-aligned and non-aligned tasks, not specifically on identity-consistent face generation with controllable attributes as in WithAnyone.

---

10. WithAnyone: Towards Controllable and ID Consistent Image Generation

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.

Appendix: Text Similarity Detection

Textual similarity detection checked 29 papers and found 3 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. WithAnyone: Towards Controllable and ID Consistent Image Generation

Detected in: Core Task (sibling), Contribution: contribution_1, Contribution: contribution_3

⚠ **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] WithAnyone: Toward Controllable and ID Consistent Image Generation View paper
- [1] Identity-preserving text-to-video generation by frequency decomposition View paper
- [2] Identity-aware infrared person image generation and re-identification via controllable diffusion model View paper
- [3] ID-Booth: Identity-consistent Face Generation with Diffusion Models View paper
- [4] Motion-I2V: Consistent and Controllable Image-to-Video Generation with Explicit Motion Modeling View paper
- [5] Animate Anyone: Consistent and Controllable Image-to-Video Synthesis for Character Animation View paper
- [6] Blip-diffusion: Pre-trained subject representation for controllable text-to-image generation and editing View paper
- [7] Controllable Human Image Generation with Personalized Multi-Garments View paper
- [8] ID-Consistent, Precise Expression Generation with Blendshape-Guided Diffusion View paper
- [9] LayerCraft: Enhancing Text-to-Image Generation with CoT Reasoning and Layered Object Integration View paper
- [10] One-Prompt-One-Story: Free-Lunch Consistent Text-to-Image Generation Using a Single Prompt View paper
- [11] Disentangled and Controllable Face Image Generation via 3D Imitative-Contrastive Learning View paper
- [12] MasaCtrl: Tuning-Free Mutual Self-Attention Control for Consistent Image Synthesis and Editing View paper
- [13] Consistent Human Image and Video Generation with Spatially Conditioned Diffusion View paper
- [14] ComposeMe: Attribute-Specific Image Prompts for Controllable Human Image Generation View paper
- [15] Tedigan: Text-guided diverse face image generation and manipulation View paper
- [16] Motioncharacter: Identity-preserving and motion controllable human video generation View paper
- [17] CharaConsist: Fine-Grained Consistent Character Generation View paper
- [18] Training-Free Consistent Text-to-Image Generation View paper

- [19] GANDiffFace: Controllable Generation of Synthetic Datasets for Face Recognition with Realistic Variations View paper
- [20] MagicNaming: Consistent Identity Generation by Finding a "Name Space" in T2I Diffusion Models View paper
- [21] Identityâ□consistent transfer learning of portraits for digital apparel sample display View paper
- [22] From parts to whole: A unified reference framework for controllable human image generation View paper
- [23] Idiff-face: Synthetic-based face recognition through fizzy identity-conditioned diffusion model View paper
- [24] Disentangled representation learning for controllable person image generation View paper
- [25] Unified Conditional Image Generation for Visible-Infrared Person Re-Identification View paper
- [26] DiffTV: Identity-preserved thermal-to-visible face translation via feature alignment and dual-stage conditions View paper
- [27] FlexIP: Dynamic Control of Preservation and Personality for Customized Image Generation View paper
- [28] CP-EB: Talking Face Generation with Controllable Pose and Eye Blinking Embedding View paper
- [29] Concat-ID: Towards Universal Identity-Preserving Video Synthesis View paper
- [30] Editable Image Elements for Controllable Synthesis View paper
- [31] Omni-id: Holistic identity representation designed for generative tasks View paper
- [32] ID: Identity-Preserving-yet-Diversified Diffusion Models for Synthetic Face Recognition View paper
- [33] Universal Fingerprint Generation: Controllable Diffusion Model With Multimodal Conditions View paper
- [34] SINGAPO: Single Image Controlled Generation of Articulated Parts in Objects View paper
- [35] Disco: Reinforcement with diversity constraints for multi-human generation View paper
- [36] Towards open-set identity preserving face synthesis View paper
- [37] PositionIC: Unified Position and Identity Consistency for Image Customization View paper
- [38] PortraitBooth: A Versatile Portrait Model for Fast Identity-Preserved Personalization View paper
- [39] Taming Encoder for Zero Fine-tuning Image Customization with Text-to-Image Diffusion Models View paper
- [40] FLUXSynID: A Framework for Identity-Controlled Synthetic Face Generation with Document and Live Images View paper
- [41] 3D Cartoon Face Generation with Controllable Expressions from a Single GAN Image View paper
- [42] AutoStory: Generating Diverse Storytelling Images with Minimal Human Efforts View paper
- [43] Linearly controllable gan: Unsupervised feature categorization and decomposition for image generation and manipulation View paper
- [44] FaceChain-FACT: Face Adapter with Decoupled Training for Identity-preserved Personalization View paper
- [45] Frame In-N-Out: Unbounded Controllable Image-to-Video Generation View paper
- [46] WithAnyone: Towards Controllable and ID Consistent Image Generation View paper
- [47] StoryMaker: Towards Holistic Consistent Characters in Text-to-image Generation View paper
- [48] Controlling Avatar Diffusion with Learnable Gaussian Embedding View paper
- [49] Face Identity-Aware Disentanglement in StyleGAN View paper
- [50] MagicPose: Realistic Human Poses and Facial Expressions Retargeting with Identity-aware Diffusion View paper
- [51] ContextGen: Contextual Layout Anchoring for Identity-Consistent Multi-Instance Generation View paper
- [52] DanceTogether! Identity-Preserving Multi-Person Interactive Video Generation View paper
- [53] Trackformer: Multi-object tracking with transformers View paper
- [54] MultiHuman-Testbench: Benchmarking Image Generation for Multiple Humans View paper
- [55] CustomVideo: Customizing Text-to-Video Generation with Multiple Subjects View paper
- [56] Diffusion Self-Distillation for Zero-Shot Customized Image Generation View paper
- [57] Id-patch: Robust id association for group photo personalization View paper
- [58] Pose-disentangled Contrastive Learning for Self-supervised Facial Representation View paper
- [59] Sample-cohesive pose-aware contrastive facial representation learning View paper
- [60] Cross-Age Contrastive Learning for Age-Invariant Face Recognition View paper
- [61] Audio-driven emotion-aware 3d talking face generation from single image View paper
- [62] Real-time, high-fidelity face identity swapping with a vision foundation model View paper
- [63] Contrastive viewpoint-aware shape learning for long-term person re-identification View paper
- [64] Face Swapping via Reverse Contrastive Learning and Explicit Identity-Attribute Disentanglement View paper
- [65] Identity-aware convolutional neural network for facial expression recognition View paper
- [66] Supervised contrastive learning with identity-label embeddings for facial action unit recognition View paper
- [67] Ominicontrol: Minimal and universal control for diffusion transformer View paper
- [68] Uniportrait: A unified framework for identity-preserving single-and multi-human image personalization View paper
- [69] SmartAvatar: Text- and Image-Guided Human Avatar Generation with VLM AI Agents View paper