Novelty Assessment Report

Paper: Referring Layer Decomposition

PDF URL: https://openreview.net/pdf?id=AlgRVfd1z7

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2025-12-30

Abstract

Precise, object-aware control over visual content is essential for advanced image editing and compositional generation. Yet, most existing approaches operate on entire images holistically, limiting the ability to isolate and manipulate individual scene elements. In contrast, layered representations, where scenes are explicitly separated into objects, environmental context, and visual effects, provide a more intuitive and structured framework for interpreting and editing visual content. To bridge this gap and enable both compositional understanding and controllable editing, we introduce the Referring Layer Decomposition (RLD) task, which predicts complete RGBA layers from a single RGB image, conditioned on flexible user prompts, such as spatial inputs (e.g., points, boxes, masks), natural language descriptions, or combinations thereof. At the core is the RefLade, a large-scale dataset comprising 1.11M image–layer–prompt triplets produced by our scalable data engine, along with 100K manually curated, high-fidelity layers. Coupled with a perceptually grounded, human-preference-aligned automatic evaluation protocol, RefLade establishes RLD as a well-defined and benchmarkable research task. Building on this foundation, we present RefLayer, a simple baseline designed for prompt-conditioned layer decomposition, achieving high visual fidelity and semantic alignment. Extensive experiments show our approach enables effective training, reliable evaluation, and high-quality image decomposition, while exhibiting strong zero-shot generalization capabilities. We will release our dataset, evaluation tools, and model for future research.

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: **Prompt-conditioned RGBA Layer Decomposition from Images**

A total of **16 papers** were analyzed and organized into a taxonomy with **11 categories**.

Taxonomy Overview

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

- **Layer Decomposition and Extraction**
- **Layer-Aware Generation**
- **Domain-Specific Layer-Based Generation**
- **Foundational Scene Understanding**

Complete Taxonomy Tree

- Prompt-conditioned RGBA Layer Decomposition from Images Survey Taxonomy
- Layer Decomposition and Extraction
  - Text-Guided Image Layer Decomposition ★ (3 papers)
  - [0] Referring Layer Decomposition (Anon et al., 2026) View paper
  - [13] Text-Guided Portrait Image Matting (Yong Xu, 2024) View paper
  - [14] Decomposition of Graphic Design with Unified Multimodal Model (H Nie, n.d.) View paper
  - Video Layer Decomposition (1 papers)
  - [1] Text2live: Text-driven layered image and video editing (Omer Bar-Tal, 2022) View paper
  - Document and Design Layer Decomposition (1 papers)
  - [12] OmniPSD: Layered PSD Generation with Diffusion Transformer (Cheng Liu, 2025) View paper
- Layer-Aware Generation
  - Text-to-RGBA Image Generation (1 papers)
  - [10] Alfie: Democratising RGBA Image Generation With No $$$ (Pippi, 2024) View paper
  - Text-to-RGBA Video Generation (2 papers)
  - [5] TransPixeler: Advancing Text-to-Video Generation with Transparency (Luozhou Wang, 2025) View paper
  - [11] TransAnimate: Taming Layer Diffusion to Generate RGBA Video (Chen Xuewei, 2025) View paper
  - Compositional Multi-Layer Scene Generation (2 papers)
  - [3] Layerflow: A unified model for layer-aware video generation (Sihui Ji, 2025) View paper
  - [9] Generating Compositional Scenes via Text-to-image RGBA Instance Generation (Fontanella, 2024) View paper
- Domain-Specific Layer-Based Generation
  - 3D Garment and Wearable Generation (2 papers)
  - [4] Garmentdreamer: 3dgs guided garment synthesis with diverse geometry and texture details (Boqian Li, 2025) View paper
  - [8] ClotheDreamer: Text-guided garment generation with 3D gaussians: Y. Liu et al. (Y Liu, 2025) View paper
  - 3D Scene and Volumetric Representation (2 papers)
  - [2] ImmerseGen: Agent-Guided Immersive World Generation with Alpha-Textured Proxies (Yuan Jinyan, 2025) View paper
  - [6] Generative Multiplane Image (GMPI): Text to Volumetric Representation (Dewei Hu, 2024) View paper
  - Document Background and Text Integration (1 papers)
  - [7] Text-Conditioned Background Generation for Editable Multi-Layer Documents (Taewon Kang, 2025) View paper

  - Vector Graphics and SVG Generation (1 papers)
  - [15] Stylesteinsvg: Example-Guided Text-to-Svg Diffusion Models Via Vectorized Stein Score Distillation (Shan Tang, n.d.) View paper
- Foundational Scene Understanding (1 papers)
  - [16] Advancing Open World Scene Understanding: From Object Detection to Image Layer Decomposition (Fangyi, n.d.) View paper

Narrative

Core task: Prompt-conditioned RGBA layer decomposition from images. The field centers on extracting and generating compositional image layers—typically with alpha channels—guided by text or other prompts. The taxonomy reveals four main branches: Layer Decomposition and Extraction focuses on parsing existing images into editable components, often leveraging matting or segmentation techniques to isolate foreground objects or semantic regions. Layer-Aware Generation emphasizes synthesizing new layered content from scratch, producing RGBA outputs that can be composited flexibly. Domain-Specific Layer-Based Generation tailors these ideas to specialized contexts such as garment modeling or document layout, where domain priors guide layer structure. Foundational Scene Understanding underpins many of these methods by providing robust representations of depth, occlusion, and semantic boundaries. Together, these branches reflect a shift from monolithic image editing toward modular, compositional workflows that afford fine-grained control.

Within Layer Decomposition and Extraction, a small handful of works explore text-guided decomposition, where natural language queries specify which elements to isolate. Referring Layer Decomposition[0] exemplifies this direction by enabling users to extract arbitrary layers via referring expressions, contrasting with earlier matting approaches like Portrait Image Matting[13] that target predefined categories. Nearby, Graphic Design Decomposition[14] tackles structured layouts rather than photographic scenes, highlighting the diversity of decomposition targets. Meanwhile, Layer-Aware Generation includes methods such as RGBA Instance Generation[9] and ImmerseGen[2], which synthesize layered assets directly, and Layerflow[3], which orchestrates multi-layer generation pipelines. The interplay between decomposition and generation remains an open question: whether to parse real images into layers or to generate layered content ab initio depends on the application, and hybrid approaches that refine extracted layers with generative priors are emerging as a promising middle ground.

Related Works in Same Category

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

1. Text-Guided Portrait Image Matting

Authors: Yong Xu, Xin Yao, Baoling Liu, Yuhui Quan, Hui Ji | **Year/Venue**: 2024 | **URL**: View paper

Abstract

Image matting is a technique used to separate the foreground of an image from the background, which estimates an alpha matte that indicates pixel-wise degree of transparency. To precisely extract target objects and address the ambiguity of solutions in image matting, many existing approaches employ a trimap or background image provided by the user as additional input to guide the matting process. This article introduces a novel matting paradigm termed text-guided image matting, utilizing a textu...

Relationship Analysis

Both papers belong to the Text-Guided Image Layer Decomposition category, using natural language prompts to guide the extraction of visual layers from images. They overlap in their use of text descriptions to specify target objects and their goal of producing RGBA outputs with transparency information. However, the original paper (Referring Layer Decomposition) focuses on multi-modal prompting (spatial + text) for general object decomposition with complete occlusion recovery, while the candidate paper specifically targets portrait matting using only text guidance without addressing occlusion completion or supporting diverse spatial prompts.

2. Decomposition of Graphic Design with Unified Multimodal Model

Authors: H Nie, Z Zhang, Y Cheng, M Yang, G Shi, et al. (6 authors total) | **URL**: View paper

Abstract

â¦ We explicitly decompose the text as a separate visual layer. (â¦ We also include text-guided, point-guided, and GPT-4 generated â¦ textual prompts for text-guided layer decomposition are â¦

Relationship Analysis

Both papers belong to the Text-Guided Image Layer Decomposition category, focusing on decomposing images into RGBA layers using natural language prompts. While the original paper (Referring Layer Decomposition) addresses general single-image decomposition with flexible multi-modal prompts (spatial and textual) for arbitrary objects and scenes, this candidate paper specifically targets graphic design posters, decomposing them into structured layers with rich metadata (font, color, alignment) and supporting interactive decomposition through text or point-based prompts. The key difference lies in domain specialization: the original paper targets general-purpose layer extraction from natural images, whereas this paper focuses on structured graphic design decomposition with design-specific metadata generation.

Contributions Analysis

Overall novelty summary. The paper introduces the Referring Layer Decomposition (RLD) task, which decomposes RGB images into RGBA layers conditioned on flexible prompts (spatial inputs, text, or combinations). According to the taxonomy, this work resides in the 'Text-Guided Image Layer Decomposition' leaf under 'Layer Decomposition and Extraction'. This leaf contains only three papers total, including the original work, indicating a relatively sparse research direction. The sibling papers focus on similar decomposition goals but differ in prompt modalities or target domains, suggesting RLD occupies a distinct niche within a small but emerging subfield.

The taxonomy reveals that 'Layer Decomposition and Extraction' sits alongside 'Layer-Aware Generation' (which synthesizes layers from scratch) and 'Domain-Specific Layer-Based Generation' (garments, 3D scenes, documents). The original paper's leaf excludes video decomposition and non-textual guidance, distinguishing it from neighboring categories like 'Video Layer Decomposition' and 'Document and Design Layer Decomposition'. The broader field comprises sixteen papers across multiple branches, with text-guided image decomposition representing a minority direction. This positioning suggests the work addresses a gap between general matting techniques and compositional generation pipelines.

Among twenty-six candidates examined, the RLD task formulation itself shows no clear refutation across six candidates, while the RefLade dataset encounters two refutable candidates among ten examined, and the evaluation protocol finds none among ten. The limited search scope—top-K semantic matches plus citation expansion—means these statistics reflect a targeted sample rather than exhaustive coverage. The dataset contribution appears to have more substantial prior work overlap, whereas the task definition and evaluation protocol seem less directly anticipated by existing literature within the examined set.

Based on the limited search of twenty-six candidates, the work appears to carve out a specific niche in prompt-conditioned layer decomposition, though the dataset component overlaps with prior efforts. The taxonomy structure confirms this is a sparsely populated research direction, with only two sibling papers in the same leaf. However, the analysis does not cover the full landscape of matting, segmentation, or compositional generation methods that may inform or constrain the novelty assessment.

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

Contribution 1: Referring Layer Decomposition (RLD) task

Description: The authors formalize a novel task that extracts targeted RGBA layers from RGB images based on multi-modal user prompts such as spatial inputs (points, boxes, masks), natural language descriptions, or combinations thereof. This task enables compositional understanding and controllable editing of visual content.

This contribution was assessed against **6 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. TransAnimate: Taming Layer Diffusion to Generate RGBA Video

URL: View paper

Brief Assessment

TransAnimate[11] focuses on generating RGBA videos with temporal coherence for game effects and visual content, not on extracting RGBA layers from single RGB images using multi-modal prompts. The tasks are fundamentally different in scope and application.

2. CreatiPoster: Towards Editable and Controllable Multi-Layer Graphic Design Generation

URL: View paper

Brief Assessment

CreatiPoster[18] focuses on generating editable multi-layer graphic designs (posters) from user prompts and assets, not on extracting RGBA layers from existing RGB images based on multi-modal prompts. The tasks address fundamentally different problems in different application domains.

3. Text2live: Text-driven layered image and video editing

URL: View paper

Brief Assessment

Text2live[1] focuses on text-driven appearance editing by generating edit layers (color+opacity) composited over original inputs, not on extracting complete RGBA layers from RGB images using multi-modal prompts. The candidate does not address the task of decomposing images into object-aware layers based on spatial or linguistic referring expressions.

4. Fine-tuning multimodal large language models for medical visual question answering: instruction tuning with region of interest attention: a thesis in Data Science

URL: View paper

Brief Assessment

Medical VQA ROI[19] focuses on medical visual question answering with region-of-interest attention for diagnostic queries, not on extracting RGBA layers from RGB images for compositional editing.

5. ImmerseGen: Agent-Guided Immersive World Generation with Alpha-Textured Proxies

URL: View paper

Brief Assessment

ImmerseGen[2] focuses on generating immersive 3D VR environments using alpha-textured proxies for real-time rendering, not on extracting RGBA layers from RGB images based on multi-modal prompts. The technical approaches and application domains are fundamentally different.

6. Art: Anonymous region transformer for variable multi-layer transparent image generation

URL: View paper

Brief Assessment

ART[17] focuses on multi-layer transparent image generation using anonymous region layouts without semantic labels, while the original paper's RLD task extracts targeted RGBA layers using multi-modal prompts (spatial inputs, natural language, or combinations). These are distinct task formulations with different input modalities and objectives.

Contribution 2: RefLade dataset and data engine

Description: The authors develop a scalable, automated data engine and use it to construct RefLade, a dataset of 1.11 million image-layer-prompt triplets with human-curated splits. This dataset establishes RLD as a trainable and benchmarkable research task.

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. MULAN: A Multi Layer Annotated Dataset for Controllable Text-to-Image Generation

URL: View paper

Prior Art Analysis

MULAN[25] demonstrates that a large-scale dataset for image layer decomposition with an automated data engine was developed prior to the original paper. MULAN[25] presents a dataset of 44,860 multi-layer annotations comprising over 101,269 instances, constructed using a scalable, automated pipeline that decomposes RGB images into RGBA layers. The data engine in MULAN[25] includes automated modules for instance discovery, completion, and re-assembly, establishing layer decomposition as a trainable task before the original paper's submission.

Evidence

Evidence 1 - **Rationale**: Both papers describe multi-stage automated pipelines for layer decomposition. MULAN[25]'s three-module approach (discovery, completion, re-assembly) parallels the original paper's six-stage engine, showing that automated data engines for this task existed prior to the original submission. - **Original**: the engine consists of six sequential stages that transform raw natural images into prompt-aligned rgba layers with complete visual content and semantic grounding, illustrated in fig. 2: (1) pre-filter: screens raw images to ensure they are suitable for decomposition based on quality, content, and o... - **Candidate**: our decomposition process into three submodules, focusing on 1) instance discovery, ordering and extraction, 2) instance completion of occluded appearance, and 3) image re-assembly as an rgba stack. each submodule is carefully designed to ensure general applicability, high instance and background re...

Evidence 2 - **Rationale**: Both papers demonstrate the application of their automated data engines to create large-scale layer decomposition datasets from public image sources, with MULAN[25] establishing this approach earlier. - **Original**: reflade consists of 430k images annotated with rgba layers, sourced from four large-scale, publicly available datasets: open images, sa-1b, coco, and objects365. on average, each image contributes 2.1 filter-passed foreground instance layers and 0.57 background layers, yielding a total of approximat... - **Candidate**: we process images from the coco [28] and laion aesthetics 6.5 [44] datasets using our novel pipeline, yielding multi-layer instance annotations for over 44k images and over 100k instances.

2. Self-supervised intrinsic image decomposition

URL: View paper

Brief Assessment

The candidate paper (Intrinsic Image Decomposition[24]) focuses on decomposing images into reflectance, shape, and lighting using self-supervised learning, not on building large-scale datasets with automated data engines for layer decomposition. The tasks and contributions are fundamentally different.

3. Cart: Compositional auto-regressive transformer for image generation

URL: View paper

Brief Assessment

CART[23] focuses on compositional autoregressive image generation through base-detail decomposition, not on layer decomposition datasets or data engines for extracting RGBA layers from images.

4. RemoteSAM: Towards Segment Anything for Earth Observation

URL: View paper

Brief Assessment

RemoteSAM[29] focuses on remote sensing/earth observation with a referring expression segmentation dataset (RemoteSAM-270k) for satellite imagery, not general image layer decomposition. The data engine and task formulation differ fundamentally from RefLade's RGBA layer decomposition approach.

5. Generative Image Layer Decomposition with Visual Effects

URL: View paper

Prior Art Analysis

Visual Effects Decomposition[20] demonstrates that a scalable, automated data engine for image layer decomposition with large-scale training data was developed prior to the original paper. The candidate paper presents a comprehensive data preparation pipeline that automatically generates simulated multi-layer data with synthesized visual effects, combining 100k foreground object images with 5m stock images to create training triplets. This automated approach to dataset construction, which includes pre-filtering, scene understanding, layer completion, and quality control stages, establishes that the concept of an automated data engine for layer decomposition datasets existed before the original work.

Evidence

Evidence 1 - **Rationale**: Both papers describe large-scale datasets with automated generation and manual curation components. The candidate's dataset preparation demonstrates prior work on automated data engines for layer decomposition. - **Original**: we presentreflade, a dataset of 1.11m image-layer-prompt triplets, including 1m auto-generated training examples, 100k manually cleaned layers, and a 10k curated test set - **Candidate**: For the simulated dataset, we use 100k foreground object images with precomputed shadows on plain backgrounds, selecting unoccluded main objects using segmentation and depth heuristics. For backgrounds, we source 5m stock images and blend the foregrounds based on object size and placement. Our camer...

Evidence 2 - **Rationale**: Both papers describe automated data engines that generate RGBA layers with quality control mechanisms. The candidate demonstrates prior work on scalable data generation pipelines for layer decomposition. - **Original**: reflade is constructed using a scalable data engine that integrates prompt interpretation, rgba synthesis, and automated filtering, which ensures both quality and extensibility for future data expansion - **Candidate**: to effectively train l ayer decomp , we curated a hybrid dataset that combines simulated and real-world data. ideally, training l ayer decomp requires image triplets: an input image irgb comp, a transparent foreground layer containing the target object and its visual effects irgba fg , and a backgro...

6. AFD-StackGAN: Automatic mask generation network for face de-occlusion using StackGAN

URL: View paper

Brief Assessment

AFD-StackGAN[22] focuses on face de-occlusion using synthetic face-occluded datasets from CelebA, not on general image layer decomposition with automated data engines for diverse scenes and objects.

7. HDR Image Generation via Gain Map Decomposed Diffusion

URL: View paper

Brief Assessment

Gain Map HDR[21] focuses on HDR image generation using gain map decomposition and constructs text-SDR-GM triplets for HDR synthesis. This differs fundamentally from RefLade's image layer decomposition task, which extracts RGBA layers from RGB images for compositional editing.

8. Cgintrinsics: Better intrinsic image decomposition through physically-based rendering

URL: View paper

Brief Assessment

CGIntrinsics[28] focuses on intrinsic image decomposition (separating images into reflectance and shading components) using physically-based rendering, not on layer decomposition with RGBA outputs and multi-modal prompting as in the original paper's RefLade contribution.

9. Learning to see through obstructions with layered decomposition

URL: View paper

Brief Assessment

Layered Decomposition Obstructions[27] focuses on removing obstructions (reflections, fences, raindrops) from image sequences using optical flow and layer reconstruction, not on creating large-scale datasets with automated data engines for general layer decomposition tasks. The candidate addresses a different problem domain (obstruction removal) rather than establishing a benchmark dataset for referring layer decomposition.

10. A generic deep architecture for single image reflection removal and image smoothing

URL: View paper

Brief Assessment

Reflection Removal Architecture[26] addresses single-image reflection removal using a cascaded CNN architecture, not large-scale dataset construction with automated data engines for general image layer decomposition. The candidate's data synthesis method (Figure 3) is limited to reflection-specific image pairs, not the 1.11M multi-modal triplets with human curation described in the original paper.

Contribution 3: Human-preference-aligned evaluation protocol

Description: The authors design an automatic evaluation protocol that assesses layer decomposition along three dimensions (preservation, completion, faithfulness) and aggregates them into a unified HPA score that strongly correlates with human judgments, enabling reliable benchmarking without human-in-the-loop evaluation.

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. Human Preference-Aligned Concept Customization Benchmark via Decomposed Evaluation

URL: View paper

Brief Assessment

Concept Customization Benchmark[35] focuses on evaluating concept customization in text-to-image generation (assessing fidelity to concept images and prompts), not layer decomposition quality assessment. The evaluation dimensions and task contexts are fundamentally different.

2. Stable Preference: Redefining Training Paradigm of Human Preference Model for Text-to-Image Synthesis

URL: View paper

Brief Assessment

Stable Preference[38] focuses on training paradigms for human preference models in text-to-image synthesis, not on evaluation protocols for image layer decomposition. The candidate addresses preference model training methodology, while the original paper proposes an evaluation protocol for assessing layer decomposition quality along preservation, completion, and faithfulness dimensions.

3. Analyzing the effect of human alignment using Bilinear Layer-wise relevance propagation

URL: View paper

Brief Assessment

Bilinear Relevance Propagation[36] focuses on analyzing neural network alignment through layer-wise relevance propagation for similarity scores, not on developing evaluation protocols for image layer decomposition quality assessment.

4. A perception-driven hybrid decomposition for multi-layer accommodative displays

URL: View paper

Brief Assessment

Accommodative Displays Decomposition[33] focuses on layer decomposition for multi-focal displays with accommodation cues, not general image layer decomposition. Their evaluation uses SSIM calibrated for specific display artifacts, not a general human-preference-aligned protocol for image quality assessment.

5. HPSv3: Towards Wide-Spectrum Human Preference Score

URL: View paper

Brief Assessment

HPSv3[34] focuses on evaluating text-to-image generation models through human preference scoring, not layer decomposition quality assessment. The evaluation dimensions and task contexts are fundamentally different.

6. Modeling human decomposition: A Bayesian approach.

URL: View paper

Brief Assessment

Human Decomposition Bayesian[37] focuses on modeling human decomposition processes using Bayesian methods, which is unrelated to image layer decomposition quality assessment or evaluation protocols for computer vision tasks.

7. iCAM06: A refined image appearance model for HDR image rendering

URL: View paper

Brief Assessment

iCAM06[32] focuses on HDR image rendering quality assessment through psychophysical experiments with human observers, not on automatic evaluation protocols for layer decomposition. The paper evaluates tone-mapping operators using paired comparisons and Thurstone scaling, which is fundamentally different from the original paper's automatic protocol for assessing layer decomposition quality along preservation, completion, and faithfulness dimensions.

8. Generative Image Layer Decomposition with Visual Effects

URL: View paper

Brief Assessment

Visual Effects Decomposition[20] uses standard metrics (PSNR, LPIPS, FID, CLIP-FID) and user studies for evaluation, but does not propose a unified human-preference-aligned automatic evaluation protocol with three dimensions (preservation, completion, faithfulness) aggregated into a single HPA score as in the original paper.

9. HALO: Human-Aligned End-to-end Image Retargeting with Layered Transformations

URL: View paper

Brief Assessment

HALO[31] focuses on image retargeting quality assessment using perceptual metrics (CLIP, DreamSim, MUSIQ, VILA) and user studies, not on layer decomposition evaluation. The candidate does not address the three-dimensional evaluation protocol (preservation, completion, faithfulness) or the HPA score aggregation method proposed in the original paper.

10. Decompose and leverage preferences from expert models for improving trustworthiness of mllms

URL: View paper

Brief Assessment

Decompose Leverage Preferences[30] focuses on evaluating MLLM responses through decomposition into atomic verification tasks for preference learning, not on evaluating image layer decomposition quality. The evaluation targets and methodologies are fundamentally different.

Appendix: Text Similarity Detection

Textual similarity detection checked 27 papers and found 2 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. MULAN: A Multi Layer Annotated Dataset for Controllable Text-to-Image Generation

Detected in: Contribution: 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] Referring Layer Decomposition View paper
- [1] Text2live: Text-driven layered image and video editing View paper
- [2] ImmerseGen: Agent-Guided Immersive World Generation with Alpha-Textured Proxies View paper
- [3] Layerflow: A unified model for layer-aware video generation View paper
- [4] Garmentdreamer: 3dgs guided garment synthesis with diverse geometry and texture details View paper
- [5] TransPixeler: Advancing Text-to-Video Generation with Transparency View paper
- [6] Generative Multiplane Image (GMPI): Text to Volumetric Representation View paper
- [7] Text-Conditioned Background Generation for Editable Multi-Layer Documents View paper
- [8] ClotheDreamer: Text-guided garment generation with 3D gaussians: Y. Liu et al. View paper
- [9] Generating Compositional Scenes via Text-to-image RGBA Instance Generation View paper
- [10] Alfie: Democratising RGBA Image Generation With No $$$ View paper
- [11] TransAnimate: Taming Layer Diffusion to Generate RGBA Video View paper
- [12] OmniPSD: Layered PSD Generation with Diffusion Transformer View paper
- [13] Text-Guided Portrait Image Matting View paper
- [14] Decomposition of Graphic Design with Unified Multimodal Model View paper
- [15] Stylesteinsvg: Example-Guided Text-to-Svg Diffusion Models Via Vectorized Stein Score Distillation View paper
- [16] Advancing Open World Scene Understanding: From Object Detection to Image Layer Decomposition View paper
- [17] Art: Anonymous region transformer for variable multi-layer transparent image generation View paper
- [18] CreatiPoster: Towards Editable and Controllable Multi-Layer Graphic Design Generation View paper
- [19] Fine-tuning multimodal large language models for medical visual question answering: instruction tuning with region of interest attention: a thesis in Data Science View paper
- [20] Generative Image Layer Decomposition with Visual Effects View paper
- [21] HDR Image Generation via Gain Map Decomposed Diffusion View paper
- [22] AFD-StackGAN: Automatic mask generation network for face de-occlusion using StackGAN View paper
- [23] Cart: Compositional auto-regressive transformer for image generation View paper
- [24] Self-supervised intrinsic image decomposition View paper
- [25] MULAN: A Multi Layer Annotated Dataset for Controllable Text-to-Image Generation View paper
- [26] A generic deep architecture for single image reflection removal and image smoothing View paper
- [27] Learning to see through obstructions with layered decomposition View paper
- [28] Cgintrinsics: Better intrinsic image decomposition through physically-based rendering View paper
- [29] RemoteSAM: Towards Segment Anything for Earth Observation View paper
- [30] Decompose and leverage preferences from expert models for improving trustworthiness of mllms View paper
- [31] HALO: Human-Aligned End-to-end Image Retargeting with Layered Transformations View paper
- [32] iCAM06: A refined image appearance model for HDR image rendering View paper
- [33] A perception-driven hybrid decomposition for multi-layer accommodative displays View paper
- [34] HPSv3: Towards Wide-Spectrum Human Preference Score View paper
- [35] Human Preference-Aligned Concept Customization Benchmark via Decomposed Evaluation View paper
- [36] Analyzing the effect of human alignment using Bilinear Layer-wise relevance propagation View paper
- [37] Modeling human decomposition: A Bayesian approach. View paper
- [38] Stable Preference: Redefining Training Paradigm of Human Preference Model for Text-to-Image Synthesis View paper