Highlighted Papers and YouTube Results

Paper: Towards RAW Object Detection in Diverse Conditions

Paper link: https://arxiv.org/abs/2411.15678

YouTube Result: I don't know.

Paper: NTClick: Achieving Precise Interactive Segmentation With Noise-tolerant Clicks

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Decoupled Distillation to Erase: A General Unlearning Method for Any Class-centric Tasks

Paper link: https://arxiv.org/abs/2503.23751

YouTube Result: I don't know.

Paper: Towards Zero-Shot Anomaly Detection and Reasoning with Multimodal Large Language Models

Paper link: https://arxiv.org/abs/2504.13399

YouTube Result: I don't know.

Paper: Cross-View Completion Models are Zero-shot Correspondence Estimators

Paper link: https://arxiv.org/abs/2412.09072

YouTube Result: I don't know.

Paper: PlanarSplatting: Accurate Planar Surface Reconstruction in 3 Minutes

Paper link: https://arxiv.org/abs/2412.03451

YouTube Result: I don't know.

Paper: Prior-free 3D Object Tracking

Paper link: https://arxiv.org/abs/2502.10606

YouTube Result: I don't know.

Paper: Gradient-Guided Annealing for Domain Generalization

Paper link: https://arxiv.org/abs/2502.20162

YouTube Result: I don't know.

Paper: Assessing and Learning Alignment of Unimodal Vision and Language Models

Paper link: https://arxiv.org/abs/2412.04616

YouTube Result: I don't know.

Paper: BEVDiffuser: Plug-and-Play Diffusion Model for BEV Denoising with Ground-Truth Guidance

Paper link: https://arxiv.org/abs/2502.19694

YouTube Result: I don't know.

Paper: HaWoR: World-Space Hand Motion Reconstruction from Egocentric Videos

Paper link: https://arxiv.org/abs/2501.02973

YouTube Result: I don't know.

Paper: ALIEN: Implicit Neural Representations for Human Motion Prediction under Arbitrary Latency

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: SkillMimic: Learning Basketball Interaction Skills from Demonstrations

Paper link: https://arxiv.org/abs/2408.15270

YouTube Result: I don't know.

Paper: Multitwine: Multi-Object Compositing with Text and Layout Control

Paper link: https://arxiv.org/abs/2502.05165

YouTube Result: I don't know.

Paper: You See it, You Got it: Learning 3D Creation on Pose-Free Videos at Scale

Paper link: https://arxiv.org/abs/2412.06699

YouTube Result: I don't know.

Paper: SP3D: Boosting Sparsely-Supervised 3D Object Detection via Accurate Cross-Modal Semantic Prompts

Paper link: https://arxiv.org/abs/2503.06467

YouTube Result: I don't know.

Paper: Structured 3D Latents for Scalable and Versatile 3D Generation

Paper link: https://arxiv.org/abs/2412.01506

YouTube Result: 1. **Motivation**: The study aims to enhance the generation of structured 3D content by combining sparse 3D structures with dense visual features from multiple views. This approach enables users to upload simple images or descriptions and receive detailed, editable 3D models, which can be integrated into various applications, including gaming and simulations.

2. **Novelty**: The novel aspect of this work is the introduction of structured latents (slat), which allow for the fusion of visual features into a latent grid. This enables the generation of high-quality, scalable 3D models from minimal input, moving beyond traditional limits in 3D content creation.

3. **Main Findings**: The main findings indicate that the structured 3D latents facilitate the creation of dynamic and editable 3D models that retain structural information. The generated models can capture intricate details such as geometry and textures, promoting adaptability for different applications.

4. **Video Title**: Trellis img-to-3d -- Structured 3D Latents for Scalable and Versatile 3D Generation (Paper Walkthru)

5. **Video Link**: [Watch here](https://www.youtube.com/watch?v=4FtVDKkXGb4)

Paper: Augmented Deep Contexts for Spatially Embedded Video Coding

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Learning Phase Distortion with Selective State Space Models for Video Turbulence Mitigation

Paper link: https://arxiv.org/abs/2504.02697

YouTube Result: I don't know.

Paper: SeedVR: Seeding Infinity in Diffusion Transformer Towards Generic Video Restoration

Paper link: https://arxiv.org/abs/2501.01320

YouTube Result: I don't know.

Paper: COUNTS: Benchmarking Object Detectors and Multimodal Large Language Models under Distribution Shifts

Paper link: https://arxiv.org/abs/2504.10158

YouTube Result: I don't know.

Paper: Memories of Forgotten Concepts

Paper link: https://arxiv.org/abs/2412.01207

YouTube Result: I don't know.

Paper: Revisiting MAE Pre-training for 3D Medical Image Segmentation

Paper link: https://arxiv.org/abs/2410.23132

YouTube Result: I don't know.

Paper: CH3Depth: Efficient and Flexible Depth Foundation Model with Flow Matching

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Lessons and Insights from a Unifying Study of Parameter-Efficient Fine-Tuning (PEFT) in Visual Recognition

Paper link: https://arxiv.org/abs/2409.16434

YouTube Result: I don't know.

Paper: Hyperbolic Safety-Aware Vision-Language Models

Paper link: https://arxiv.org/abs/2503.12127

YouTube Result: I don't know.

Paper: v-CLR: View-Consistent Learning for Open-World Instance Segmentation

Paper link: https://arxiv.org/abs/2504.01383

YouTube Result: I don't know.

Paper: Satellite Observations Guided Diffusion Model for Accurate Meteorological States at Arbitrary Resolution

Paper link: https://arxiv.org/abs/2502.07814

YouTube Result: I don't know.

Paper: ESC: Erasing Space Concept for Knowledge Deletion

Paper link: https://arxiv.org/abs/2504.02199

YouTube Result: I don't know.

Paper: Goku: Flow Based Video Generative Foundation Models

Paper link: https://arxiv.org/abs/2502.04896

YouTube Result: 1. **Motivation**: The motivation behind the study of Goku is to advance the field of video generation by creating a foundation model that can produce high-quality videos from text prompts, thereby demonstrating the capabilities of AI in generating creative content.

2. **Novelty**: The novel aspects of the study include the use of rectified flow Transformers, which

provide a more efficient and effective way to transform noise into realistic images and videos. This approach allows Goku to understand motion rather than just treating video as a sequence of frames.

3. **Main Findings**: Goku has achieved state-of-the-art performance in both image and video generation, scoring 84.85% in text-to-video generation benchmarks. It was trained on a massive dataset consisting of 36 million video-text pairs and 160 million image-text pairs, enabling it to generate high-quality content with remarkable accuracy.

4. **Video Title**: Goku: Flow Based Video Generative Foundation Models

5. **Video Link**: [Watch here](https://www.youtube.com/watch?v=coh1ya9Rx4A)

Paper: X-Dyna: Expressive Dynamic Human Image Animation

Paper link: https://arxiv.org/abs/2501.10021

YouTube Result: 1. **Motivation**: The study aims to animate still images of people, such as old family photos, transforming them into dynamic videos that depict movement, like dancing.

2. **Novelty**: The novelty of the study lies in its ability to transfer dance movements onto images of different individuals and animate the background in a realistic manner. It employs a technique akin to zero-shot learning, allowing the AI to animate movements it hasn't been specifically trained on.

3. **Main Findings**: The main findings indicate that the method significantly improves the naturalness of movement in animations, overcoming previous limitations where only the person would move while the background remained static, leading to an unrealistic appearance.

4. **Video Title**: X Dyna Expressive Dynamic Human Image Animation

5. **Video Link**: [X Dyna Expressive Dynamic Human Image Animation](https://www.youtube.com/watch?v=qSDj7hhpZMg)

Paper: MangaNinja: Line Art Colorization with Precise Reference Following

Paper link: https://arxiv.org/abs/2501.08332

YouTube Result: 1. **Motivation**: The motivation behind the study of MangaNinja is to revolutionize the process of line art colorization, allowing anime artists to transform intricate black and white drawings into beautifully colored artworks quickly and efficiently. This addresses the time-consuming and tedious nature of traditional colorization methods.

2. **Novelty**: The novel aspects of the study include the use of a dual branch structure that incorporates advanced techniques such as diffusion models, patch shuffling, and interactive point-driven guidance. Unlike previous AI attempts, MangaNinja learns from actual anime videos, capturing the nuances of movement and color transitions, which enhances the accuracy of colorization.

3. **Main Findings**: The main findings indicate that MangaNinja can accurately color line art while matching the colors to specific reference images. It effectively manages challenges such as mismatched references and multi-character compositions. The system's performance on a custom benchmark demonstrates its efficiency and reliability, making it a game-changer for the anime industry.

4. **Video Title**: MangaNinja: Line Art Colorization with Precise Reference Following (Paper Walkthrough)

5. **Video Link**: [MangaNinja: Line Art Colorization with Precise Reference Following](https://www.youtube.com/watch?v=63ZK40J5nD0)

Paper: PartGen: Part-level 3D Generation and Reconstruction with Multi-view Diffusion Models

Paper link: https://arxiv.org/abs/2412.18608

YouTube Result: 1. **Motivation**: The study aims to provide a tool that allows users to create and edit 3D models by generating them part by part, based on simple inputs like text or images. This addresses the need for more intuitive and accessible 3D modeling, especially for those without advanced skills in 3D modeling software.

2. **Novelty**: The novel aspect of this study is the use of a diffusion-based network to segment objects into meaningful, consistent parts, which can be assembled into a complete 3D model. This capability of part-level generation and editing sets it apart from previous models that produced single-mesh outputs.

3. **Main Findings**: The findings indicate that PartGen effectively enables 3D generation from various input types, allows for part editing, and supports real-world decomposition of 3D objects. The tool is positioned as a significant advancement in the field of 3D modeling technology.

4. **Video Title**: PartGen: Part-level 3D Generation and Reconstruction with Multi-View Diffusion Models

5. **Video Link**: [PartGen: Part-level 3D Generation and Reconstruction with Multi-View Diffusion Models](https://www.youtube.com/watch?v=e9ABYNKA7tc)

Paper: DepthCrafter: Generating Consistent Long Depth Sequences for Open-world Videos

Paper link: https://arxiv.org/abs/2409.02095

YouTube Result: 1. **Motivation**: The study aims to address the challenges of estimating depth in videos, which can vary greatly in content, motion, and length. The motivation is to improve the understanding of video spatial layout by providing accurate depth information.

2. **Novelty**: The novelty of the study lies in the development of DepthCrafter, a method that utilizes a special strategy and paired video depth datasets to train an AI model, enabling the accurate generation of depth sequences in videos.

3. **Main Findings**: DepthCrafter effectively estimates depth in videos, allowing for a 3D representation of the spatial layout by assigning distance values to pixels. This leads to improved understanding of object distances and the overall scene composition in video content.

4. **Video Title**: DepthCrafter - Install Locally - Generate Depth Sequences for Open-world Videos

5. **Video Link**: https://www.youtube.com/watch?v=anZGKW4nFe4

Paper: Reference-Based 3D-Aware Image Editing with Triplanes

Paper link: https://arxiv.org/abs/2404.03632

YouTube Result: I don't know.

Paper: WonderWorld: Interactive 3D Scene Generation from a Single Image

Paper link: https://arxiv.org/abs/2406.09394

YouTube Result: I don't know.

Paper: MAtCha Gaussians: Atlas of Charts for High-Quality Geometry and Photorealism From Sparse Views

Paper link: https://arxiv.org/abs/2412.06767

YouTube Result: I don't know.

Paper: Taming Video Diffusion Prior with Scene-Grounding Guidance for 3D Gaussian Splatting from Sparse Inputs

Paper link: https://arxiv.org/abs/2503.05082

YouTube Result: I don't know.

Paper: MoSca: Dynamic Gaussian Fusion from Casual Videos via 4D Motion Scaffolds

Paper link: https://arxiv.org/abs/2405.17421

YouTube Result: 1. **Motivation**: The study aims to tackle the challenging task of reconstructing a dynamic 4D scene from unposed monocular RGB videos captured in the wild. This is particularly difficult due to the ill-posed nature of the inverse problem involved.

2. **Novelty**: The novel aspect of the study is the introduction of the MoSca representation, which utilizes a sparse graph of SE3 motion trajectories. This representation enables a compact and smooth encoding of the underlying motion, allowing for interpolation into a dense SE3 deformation field.

3. **Main Findings**: The study presents visual results from diverse casual videos and demonstrates comparisons on various benchmarks, including DIE and Nvidia benchmarks. The method effectively reconstructs and renders dynamic scenes from monocular video inputs.

4. **Video Title**: MoSca-Version2: Dynamic Gaussian Fusion from Casual Videos via 4D Motion Scaffolds

5. **Video Link**: [Watch the video](https://www.youtube.com/watch?v=7WrG5-xH1_k)

Paper: DualPM: Dual Posed-Canonical Point Maps for 3D Shape and Pose Reconstruction

Paper link: https://arxiv.org/abs/2412.04464

YouTube Result: I don't know.

Paper: EBS-EKF: Accurate and High Frequency Event-based Star Tracking

Paper link: https://arxiv.org/abs/2503.20101

YouTube Result: I don't know.

Paper: RoboPEPP: Vision-Based Robot Pose and Joint Angle Estimation through Embedding Predictive Pre-Training

Paper link: https://arxiv.org/abs/2411.17662

YouTube Result: 1. **Motivation**: The study aims to improve the estimation of joint angles and robot pose from a single image, which is crucial for applications such as human-robot interaction and multi-robot collaboration.

2. **Novelty**: The novelty of the study lies in the introduction of Robep, a framework that integrates the robot's physical model into an encoder predictor network using a masking-based embedding predictive pre-training approach, enhancing the system's ability to handle occlusions and truncations.

3. **Main Findings**: Robep demonstrates improved robustness against occlusions and partial visibility by fine-tuning the network with joint and keypoint prediction, and employing a keypoint-based filtering method. The embedding predictive pre-training allows the model to better infer joint information from the surrounding image context.

4. **Video Title**: [CVPR 2025] RoboPEPP

5. **Video Link**: [Watch here](https://www.youtube.com/watch?v=pbM60-kHSdE)

Paper: MonSter: Marry Monodepth to Stereo Unleashes Power

Paper link: https://arxiv.org/abs/2501.08643

YouTube Result: 1. **Motivation**: The study aims to tackle the challenge of accurately estimating depth from images, particularly addressing issues faced by existing stereo matching methods such as occlusions and textureless surfaces.

2. **Novelty**: The novel aspect of the study is the introduction of a dual branch architecture that combines the strengths of monocular depth estimation and stereo matching, allowing these two approaches to support each other iteratively.

3. **Main Findings**: The results show that the proposed method, MonSter, achieves strong performance, ranking first on several major benchmarks and demonstrating zero-shot generalization. It effectively produces cleaner and more accurate depth maps in difficult areas compared to baseline methods, even when trained on different datasets.

4. **Video Title**: MonSter: Better Depth with Mono & Stereo Vision

5. **Video Link**: [Watch here](https://www.youtube.com/watch?v=65ZSR9AY8EM)

Paper: InteractAnything: Zero-shot Human Object Interaction Synthesis via LLM Feedback and Object Affordance Parsing

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Image Quality Assessment: From Human to Machine Preference

Paper link: https://arxiv.org/abs/2503.10078

YouTube Result: I don't know.

Paper: Multirate Neural Image Compression with Adaptive Lattice Vector Quantization

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: OpenHumanVid: A Large-Scale High-Quality Dataset for Enhancing Human-Centric Video Generation

Paper link: https://arxiv.org/abs/2412.00115

YouTube Result: I don't know.

Paper: Generative Photography: Scene-Consistent Camera Control for Realistic Text-to-Image Synthesis

Paper link: https://arxiv.org/abs/2412.02168

YouTube Result: I don't know.

Paper: SoundVista: Novel-View Ambient Sound Synthesis via Visual-Acoustic Binding

Paper link: https://arxiv.org/abs/2504.05576

YouTube Result: I don't know.

Paper: MambaVLT: Time-Evolving Multimodal State Space Model for Vision-Language Tracking

Paper link: https://arxiv.org/abs/2411.15459

YouTube Result: I don't know.

Paper: Theoretical Insights in Model Inversion Robustness and Conditional Entropy Maximization for Collaborative Inference Systems

Paper link: https://arxiv.org/abs/2503.00383

YouTube Result: I don't know.

Paper: Modeling Thousands of Human Annotators for Generalizable Text-to-Image Person Re-identification

Paper link: https://arxiv.org/abs/2503.09962

YouTube Result: I don't know.

Paper: Graph Neural Network Combining Event Stream and Periodic Aggregation for Low-Latency Event-based Vision

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Enhanced Visual-Semantic Interaction with Tailored Prompts for Pedestrian Attribute Recognition

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Timestep Embedding Tells: It's Time to Cache for Video Diffusion Model

Paper link: https://arxiv.org/abs/2411.19108

YouTube Result: I don't know.

Paper: Scaling Vision Pre-Training to 4K Resolution

Paper link: https://arxiv.org/abs/2503.19903

YouTube Result: 1. **Motivation**: The motivation behind the study is to enhance visual perception and detail in various applications, such as medical imaging and digital mapping, by pre-training vision models at higher resolutions.

2. **Novelty**: The novel aspect of the study is the introduction of scale selective processing (PS3), which allows for effective pre-training of vision models at 4K resolution without significantly increasing computational demands.

3. **Main Findings**: The main finding is that PS3 enables the pre-training of vision models at a stunning 4K resolution, overcoming challenges that have previously made this difficult.

4. **Video Title**: Scaling Vision Pre-Training to 4K Resolution

5. **Video Link**: [Watch here](https://www.youtube.com/watch?v=hGyIc_wG2yw)

Paper: Multimodal Autoregressive Pre-training of Large Vision Encoders

Paper link: https://arxiv.org/abs/2411.14402

YouTube Result: 1. **Motivation**: The study is motivated by the need to create AI systems that can simultaneously understand images and text, moving beyond traditional models that specialize in only one of these modalities.

2. **Novelty**: The novel aspect of the study is the introduction of a multimodal autoregressive pre-training approach, where the AI learns to predict the next part of an image or the next word in a sentence, similar to how humans anticipate outcomes based on context.

3. **Main Findings**: The findings indicate that the new AI model, AIMV2, can outperform existing models in tasks such as identifying objects and understanding relationships within images. This model's learning approach mimics human learning, making it more effective at processing multimodal information.

4. **Video Title**: Multimodal Autoregressive Pre-training of Large Vision Encoders

5. **Video Link**: [Watch here](https://www.youtube.com/watch?v=qzpz_DKIWjM)

Paper: Learning Class Prototypes for Unified Sparse-Supervised 3D Object Detection

Paper link: https://arxiv.org/abs/2503.21099

YouTube Result: I don't know.

Paper: Open-Canopy: Towards Very High Resolution Forest Monitoring

Paper link: https://arxiv.org/abs/2407.09392

YouTube Result: I don't know.

Paper: HOT3D: Hand and Object Tracking in 3D from Egocentric Multi-View Videos

Paper link: https://arxiv.org/abs/2411.19167

YouTube Result: I don't know.

Paper: Doppelgängers and Adversarial Vulnerability

Paper link: https://arxiv.org/abs/2410.13193

YouTube Result: I don't know.

Paper: Annotation Ambiguity Aware Semi-Supervised Medical Image Segmentation

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: EffiDec3D: An Optimized Decoder for High-Performance and Efficient 3D Medical Image Segmentation

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: HotSpot: Signed Distance Function Optimization with an Asymptotically Sufficient Condition

Paper link: https://arxiv.org/abs/2411.14628

YouTube Result: I don't know.

Paper: Hardware-Rasterized Ray-Based Gaussian Splatting

Paper link: https://arxiv.org/abs/2503.18682

YouTube Result: I don't know.

Paper: OpticalNet: An Optical Imaging Dataset and Benchmark Beyond the Diffraction Limit

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Scaling Inference Time Compute for Diffusion Models

Paper link: https://arxiv.org/abs/2505.01823

YouTube Result: I don't know.

Paper: MAR-3D: Progressive Masked Auto-regressor for High-Resolution 3D Generation

Paper link: https://arxiv.org/abs/2503.20519

YouTube Result: I don't know.

Paper: DashGaussian: Optimizing 3D Gaussian Splatting in 200 Seconds

Paper link: https://arxiv.org/abs/2503.18402

YouTube Result: I don't know.

Paper: Matrix3D: Large Photogrammetry Model All-in-One

Paper link: https://arxiv.org/abs/2502.07685

YouTube Result: I don't know.

Paper: DroneSplat: 3D Gaussian Splatting for Robust 3D Reconstruction from In-the-Wild Drone Imagery

Paper link: https://arxiv.org/abs/2503.16964

YouTube Result: I don't know.

Paper: NeRFPrior: Learning Neural Radiance Field as a Prior for Indoor Scene Reconstruction

Paper link: https://arxiv.org/abs/2503.18361

YouTube Result: I don't know.

Paper: QuCOOP: A Versatile Framework for Solving Composite and Binary-Parametrised Problems on Quantum Annealers

Paper link: https://arxiv.org/abs/2503.19718

YouTube Result: I don't know.

Paper: Image Reconstruction from Readout-Multiplexed Single-Photon Detector Arrays

Paper link: https://arxiv.org/abs/2312.02971

YouTube Result: I don't know.

Paper: Learning to Filter Outlier Edges in Global SfM

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Structure-Aware Correspondence Learning for Relative Pose Estimation

Paper link: https://arxiv.org/abs/2503.18671

YouTube Result: I don't know.

Paper: CRISP: Object Pose and Shape Estimation with Test-Time Adaptation

Paper link: https://arxiv.org/abs/2412.01052

YouTube Result: I don't know.

Paper: CAP-Net: A Unified Network for 6D Pose and Size Estimation of Categorical Articulated Parts from a Single RGB-D Image

Paper link: https://arxiv.org/abs/2504.11230

YouTube Result: I don't know.

Paper: Tuning the Frequencies: Robust Training for Sinusoidal Neural Networks

Paper link: https://arxiv.org/abs/2407.21121

YouTube Result: I don't know.

Paper: Detection-Friendly Nonuniformity Correction: A Union Framework for Infrared UAV Target Detection

Paper link: https://arxiv.org/abs/2504.04012

YouTube Result: I don't know.

Paper: SimLingo: Vision-Only Closed-Loop Autonomous Driving with Language-Action Alignment

Paper link: https://arxiv.org/abs/2503.09594

YouTube Result: I don't know.

Paper: HSI-GPT: A General-Purpose Large Scene-Motion-Language Model for Human Scene Interaction

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: DiffusionDrive: Truncated Diffusion Model for End-to-End Autonomous Driving

Paper link: https://arxiv.org/abs/2411.15139

YouTube Result: I don't know.

Paper: MPDrive: Improving Spatial Understanding with Marker-Based Prompt Learning for Autonomous Driving

Paper link: https://arxiv.org/abs/2504.00379

YouTube Result: I don't know.

Paper: Reasoning in Visual Navigation of End-to-end Trained Agents: A Dynamical Systems Approach

Paper link: https://arxiv.org/abs/2503.08306

YouTube Result: I don't know.

Paper: Seurat: From Moving Points to Depth

Paper link: https://arxiv.org/abs/2504.14687

YouTube Result: I don't know.

Paper: ManiVideo: Generating Hand-Object Manipulation Video with Dexterous and Generalizable Grasping

Paper link: https://arxiv.org/abs/2412.16212

YouTube Result: I don't know.

Paper: InterMimic: Towards Universal Whole-Body Control for Physics-Based Human-Object Interactions

Paper link: https://arxiv.org/abs/2502.20390

YouTube Result: 1. **Motivation**: The study aims to develop a framework for universal whole-body control in humanoid robots, enabling them to perform physics-based interactions with various objects. The motivation stems from the need for robots to engage in dynamic and complex human-object interactions effectively.

2. **Novelty**: The novel aspect of the study is the introduction of the InterMimic framework, which allows a single policy to handle diverse whole-body interaction skills without relying on external retargeting. This integration of retargeting into the imitation process represents a significant advancement in the field.

3. **Main Findings**: The study demonstrates that InterMimic can synthesize highly dynamic interactions and learn directly from motion capture data, resulting in smoother and more natural interactions while correcting contact penetration artifacts. The framework shows improved reliability and effectiveness in handling contact-rich interactions compared to previous methods.

4. **Video Title**: InterMimic: Towards Universal Whole-Body Control for Physics-Based Human-Object Interactions

5. **Video Link**: [Watch here](https://www.youtube.com/watch?v=ZJT387dvI9w)

Paper: ClimbingCap: Multi-Modal Dataset and Method for Rock Climbing in World Coordinate

Paper link: https://arxiv.org/abs/2503.21268

YouTube Result: I don't know.

Paper: FineVQ: Fine-Grained User Generated Content Video Quality Assessment

Paper link: https://arxiv.org/abs/2412.19238

YouTube Result: I don't know.

Paper: Volumetrically Consistent 3D Gaussian Rasterization

Paper link: https://arxiv.org/abs/2502.08297

YouTube Result: I don't know.

Paper: UniReal: Universal Image Generation and Editing via Learning Real-world Dynamics

Paper link: https://arxiv.org/abs/2412.07774

YouTube Result: 1. **Motivation**: The study aims to advance the field of image generation and editing by incorporating real-world dynamics, allowing for more realistic and versatile image manipulation.

2. **Novelty**: The novel aspect of this study is the integration of understanding real-world dynamics into the image generation process, which enhances the realism and flexibility of generated images compared to traditional methods.

3. **Main Findings**: The main findings indicate that the proposed method significantly improves the quality and applicability of generated images in various contexts, showcasing its potential for practical use in diverse applications.

4. **Video Title**: UniReal: Universal Image Generation and Editing via Learning Real-world Dynamics

5. **Video Link**: https://www.youtube.com/watch?v=zwOedmmEGv4

Paper: Extrapolating and Decoupling Image-to-Video Generation Models: Motion Modeling is Easier Than You Think

Paper link: https://arxiv.org/abs/2503.00948

YouTube Result: I don't know.

Paper: SKDream: Controllable Multi-view and 3D Generation with Arbitrary Skeletons

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: High-Fidelity Lightweight Mesh Reconstruction from Point Clouds

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Parallelized Autoregressive Visual Generation

Paper link: https://arxiv.org/abs/2504.08388

YouTube Result: 1. **Motivation**: The motivation behind the study is to significantly speed up the process of AI image and video generation, which traditionally has been slow due to the pixel-by-pixel generation method.

2. **Novelty**: The novel aspect of the study is the introduction of parallelized autoregressive generation, which allows for predicting visual elements simultaneously rather than one after the other, thereby enhancing the speed of content creation.

3. **Main Findings**: The main findings indicate that not all visual elements depend equally on each other, which allows for a more efficient generation process that can produce high-quality images and videos much faster than traditional methods.

4. **Video Title**: Parallelized Autoregressive Visual Generation

5. **Video Link**: https://www.youtube.com/watch?v=iGmhCoApmHE

Paper: Driving by the Rules: A Benchmark for Integrating Traffic Sign Regulations into Vectorized HD Map

Paper link: https://arxiv.org/abs/2410.23780

YouTube Result: I don't know.

Paper: TKG-DM: Training-free Chroma Key Content Generation Diffusion Model

Paper link: https://arxiv.org/abs/2411.15580

YouTube Result: I don't know.

Paper: SCSA: A Plug-and-Play Semantic Continuous-Sparse Attention for Arbitrary Semantic Style Transfer

Paper link: https://arxiv.org/abs/2503.04119

YouTube Result: I don't know.

Paper: Towards Explainable and Unprecedented Accuracy in Matching Challenging Finger Crease Patterns

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: From Words to Structured Visuals: A Benchmark and Framework for Text-to-Diagram Generation and Editing

Paper link: https://arxiv.org/abs/2411.11916

YouTube Result: I don't know.

Paper: Digital Twin Catalog: A Large-Scale Photorealistic 3D Object Digital Twin Dataset

Paper link: https://arxiv.org/abs/2504.08541

YouTube Result: I don't know.

Paper: FirePlace: Geometric Refinements of LLM Common Sense Reasoning for 3D Object Placement

Paper link: https://arxiv.org/abs/2503.04919

YouTube Result: I don't know.

Paper: Seeing More with Less: Human-like Representations in Vision Models

Paper link: https://arxiv.org/abs/1812.02378

YouTube Result: I don't know.

Paper: DistinctAD: Distinctive Audio Description Generation in Contexts

Paper link: https://arxiv.org/abs/2411.18180

YouTube Result: I don't know.

Paper: Deep Fair Multi-View Clustering with Attention KAN

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Unsupervised Continual Domain Shift Learning with Multi-Prototype Modeling

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: VEU-Bench: Towards Comprehensive Understanding of Video Editing

Paper link: https://arxiv.org/abs/2504.17828

YouTube Result: I don't know.

Paper: Question-Aware Gaussian Experts for Audio-Visual Question Answering

Paper link: https://arxiv.org/abs/2503.04459

YouTube Result: I don't know.

Paper: Instruction-based Image Manipulation by Watching How Things Move

Paper link: https://arxiv.org/abs/2412.12087

YouTube Result: I don't know.

Paper: SnapGen: Taming High-Resolution Text-to-Image Models for Mobile Devices with Efficient Architectures and Training

Paper link: https://arxiv.org/abs/2412.09619

YouTube Result: I don't know.

Paper: MASH-VLM: Mitigating Action-Scene Hallucination in Video-LLMs through Disentangled Spatial-Temporal Representations

Paper link: https://arxiv.org/abs/2503.15871

YouTube Result: I don't know.

Paper: UMotion: Uncertainty-driven Human Motion Estimation from Inertial and Ultra-wideband Units

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: MammAlps: A Multi-view Video Behavior Monitoring Dataset of Wild Mammals in the Swiss Alps

Paper link: https://arxiv.org/abs/2503.18223

YouTube Result: I don't know.

Paper: GroundingFace: Fine-grained Face Understanding via Pixel Grounding Multimodal Large Language Model

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Exact: Exploring Space-Time Perceptive Clues for Weakly Supervised Satellite Image Time Series Semantic Segmentation

Paper link: https://arxiv.org/abs/2412.03968

YouTube Result: I don't know.

Paper: Dr. Splat: Directly Referring 3D Gaussian Splatting via Direct Language Embedding Registration

Paper link: https://arxiv.org/abs/2502.16652

YouTube Result: I don't know.

Paper: Style Evolving along Chain-of-Thought for Unknown-Domain Object Detection

Paper link: https://arxiv.org/abs/2503.09968

YouTube Result: I don't know.

Paper: Olympus: A Universal Task Router for Computer Vision Tasks

Paper link: https://arxiv.org/abs/2412.09612

YouTube Result: 1. **Motivation**: The study aims to enhance the efficiency and flexibility of computer vision tasks by developing a universal task router that can intelligently direct various tasks to specialized tools based on user instructions.

2. **Novelty**: The innovative aspect of this study is the integration of multimodal large language models (MLMs) to create a system that can manage over 20 different vision-related tasks, including image generation, video editing, and 3D modeling, all orchestrated through a single controller.

3. **Main Findings**: Olympus demonstrates versatility by successfully executing tasks based on text instructions, such as generating images, editing existing ones, and converting descriptions into 3D models. It can also string together a series of tasks, showcasing its capability to handle complex workflows.

4. **Video Title**: Olympus: Smarter Computer Vision with Task Routing

5. **Video Link**: [Watch here](https://www.youtube.com/watch?v=Lc4iDiG-O3M)

Paper: Filter Images First, Generate Instructions Later: Pre-Instruction Data Selection for Visual Instruction Tuning

Paper link: https://arxiv.org/abs/2503.07591

YouTube Result: 1. **Motivation**: The study aims to make visual instruction tuning more accessible and efficient by addressing the resource-intensive nature of current methods, which often require massive datasets of image-instruction pairs. This high demand for data can be a barrier to entry for researchers and developers looking to utilize this technology.

2. **Novelty**: The novel aspect of this study is the proposed approach of selecting the best unlabeled images before generating instructions, which flips the traditional method on its head. Instead of beginning with a large amount of pre-labeled data, this study suggests focusing on a more curated selection of images.

3. **Main Findings**: The main findings indicate that by filtering images first and then generating instructions specifically for those images, the process of visual instruction tuning can become more efficient and less resource-intensive. This strategic selection could lower costs and streamline the workflow in developing visual language models.

4. **Video Title**: Filter Images First: Efficient Visual Instruction Tuning!

5. **Video Link**: [https://www.youtube.com/watch?v=8EePj6CbayY](https://www.youtube.com/watch?v=8EePj6Cbay

Y)

Paper: Holmes-VAU: Towards Long-term Video Anomaly Understanding at Any Granularity

Paper link: https://arxiv.org/abs/2412.06171

YouTube Result: I don't know.

Paper: Can Machines Understand Composition? Dataset and Benchmark for Photographic Image Composition Embedding and Understanding

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: MLLM-as-a-Judge for Image Safety without Human Labeling

Paper link: https://arxiv.org/abs/2501.00192

YouTube Result: I don't know.

Paper: LLMDet: Learning Strong Open-Vocabulary Object Detectors under the Supervision of Large Language Models

Paper link: https://arxiv.org/abs/2501.18954

YouTube Result: I don't know.

Paper: EventPSR: Surface Normal and Reflectance Estimation from Photometric Stereo Using an Event Camera

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: OPTICAL: Leveraging Optimal Transport for Contribution Allocation in Dataset Distillation

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Less is More: Efficient Model Merging with Binary Task Switch

Paper link: https://arxiv.org/abs/2412.00054

YouTube Result: I don't know.

Paper: KAC: Kolmogorov-Arnold Classifier for Continual Learning

Paper link: https://arxiv.org/abs/2503.21076

YouTube Result: I don't know.

Paper: Label Shift Meets Online Learning: Ensuring Consistent Adaptation with Universal Dynamic Regret

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Overcoming Shortcut Problem in VLM for Robust Out-of-Distribution Detection

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: TFCustom: Customized Image Generation with Time-Aware Frequency Feature Guidance

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: CCIN: Compositional Conflict Identification and Neutralization for Composed Image Retrieval

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Towards Improved Text-Aligned Codebook Learning: Multi-Hierarchical Codebook-Text Alignment with Long Text

Paper link: https://arxiv.org/abs/2503.01261

YouTube Result: I don't know.

Paper: Learning Conditional Space-Time Prompt Distributions for Video Class-Incremental Learning

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Task-driven Image Fusion with Learnable Fusion Loss

Paper link: https://arxiv.org/abs/2412.03240

YouTube Result: I don't know.

Paper: EmoDubber: Towards High Quality and Emotion Controllable Movie Dubbing

Paper link: https://arxiv.org/abs/2412.08988

YouTube Result: I don't know.

Paper: From Faces to Voices: Learning Hierarchical Representations for High-quality Video-to-Speech

Paper link: https://arxiv.org/abs/2503.16956

YouTube Result: I don't know.

Paper: Diffusion-based Realistic Listening Head Generation via Hybrid Motion Modeling

Paper link: Not found on arXiv

YouTube Result: 1. **Motivation**: The study aims to enhance the realism and effectiveness of head generation in listening scenarios, particularly through the use of diffusion-based methods and hybrid motion modeling techniques.

2. **Novelty**: The novel aspect of the study lies in its unique approach to combining diffusion processes with hybrid motion modeling, which allows for more realistic representations of head movements and listening postures compared to traditional methods.

3. **Main Findings**: The study found that their proposed method significantly improves the quality of head generation, making the listening experience more immersive and realistic. The results demonstrate the effectiveness of integrating diffusion-based techniques with motion modeling to

achieve these enhancements.

4. **Video Title**: Diffusion-based Realistic Listening Head Generation via Hybrid Motion Modeling Supplementary Video

5. **Video Link**: [Watch here](https://www.youtube.com/watch?v=v0pq9YqmAjw)

Paper: FreeCloth: Free-form Generation Enhances Challenging Clothed Human Modeling

Paper link: https://arxiv.org/abs/2411.19942

YouTube Result: 1. **Motivation**: The study is motivated by the need to create realistic and animatable human avatars that can be accurately modeled while wearing various garments, particularly in the context of 3D poses.

2. **Novelty**: The novel aspect of the study is the introduction of a hybrid framework that segments the human body into three distinct regions (unclothed, deformed, and generated) to effectively manage clothing modeling, utilizing both traditional deformation techniques and a free form part-aware generation module for loose clothing.

3. **Main Findings**: The approach demonstrated superior visual quality and realism in modeling clothed humans compared to previous methods, particularly in handling loose clothing. The method captures high-fidelity details and significantly improves the overall representation of clothed avatars.

4. **Video Title**: [CVPR 2025 Highlight] FreeCloth: Free-form Generation Enhances Challenging Clothed Human Modeling

5. **Video Link**: [Watch here](https://www.youtube.com/watch?v=4dyM1hjTBdQ)

Paper: Volume Tells: Dual Cycle-Consistent Diffusion for 3D Fluorescence Microscopy De-noising and Super-Resolution

Paper link: https://arxiv.org/abs/2503.02261

YouTube Result: I don't know.

Paper: UltraFusion: Ultra High Dynamic Imaging using Exposure Fusion

Paper link: https://arxiv.org/abs/2501.11515

YouTube Result: I don't know.

Paper: SpecTRe-GS: Modeling Highly Specular Surfaces with Reflected Nearby Objects by Tracing Rays in 3D Gaussian Splatting

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Inst3D-LMM: Instance-Aware 3D Scene Understanding with Multi-modal Instruction Tuning

Paper link: https://arxiv.org/abs/2503.00513

YouTube Result: I don't know.

Paper: Flowing from Words to Pixels: A Noise-Free Framework for Cross-Modality Evolution

Paper link: https://arxiv.org/abs/2412.15213

YouTube Result: I don't know.

Paper: No Pains, More Gains: Recycling Sub-Salient Patches for Efficient High-Resolution Image Recognition

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Light Transport-aware Diffusion Posterior Sampling for Single-View Reconstruction of 3D Volumes

Paper link: https://arxiv.org/abs/2501.05226

YouTube Result: I don't know.

Paper: Rethinking Personalized Aesthetics Assessment: Employing Physique Aesthetics Assessment as An Exemplification

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Breaking the Memory Barrier of Contrastive Loss via Tile-Based Strategy

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: ArcPro: Architectural Programs for Structured 3D Abstraction of Sparse Points

Paper link: https://arxiv.org/abs/2503.02745

YouTube Result: I don't know.

Paper: Your Large Vision-Language Model Only Needs A Few Attention Heads For Visual Grounding

Paper link: https://arxiv.org/abs/2503.06287

YouTube Result: I don't know.

Paper: Structure from Collision

Paper link: https://arxiv.org/abs/2505.04134

YouTube Result: 1. **Motivation**: The motivation behind the study is to understand the processes involved in repairing vehicles that have sustained structural damage due to collisions, as car accidents are very common.

2. **Novelty**: The novel aspects of the study include the use of advanced technology such as frame machines and digital measuring systems that ensure precise realignment of a vehicle's structure after a collision.

3. **Main Findings**: The main findings reveal that collision centers can effectively repair vehicles by

utilizing specialized tools and methods to restore structural integrity and that digital technology plays a crucial role in identifying and reversing damage.

4. **Video Title**: How Wrecked Cars Are Repaired | Cars Insider

5. **Video Link**: [How Wrecked Cars Are Repaired](https://www.youtube.com/watch?v=mYmNM8-XRP0)

Paper: OmniSplat: Taming Feed-Forward 3D Gaussian Splatting for Omnidirectional Images with Editable Capabilities

Paper link: https://arxiv.org/abs/2412.16604

YouTube Result: I don't know.

Paper: VideoScene: Distilling Video Diffusion Model to Generate 3D Scenes in One Step

Paper link: https://arxiv.org/abs/2504.01956

YouTube Result: I don't know.

Paper: USP-Gaussian: Unifying Spike-based Image Reconstruction, Pose Correction and Gaussian Splatting

Paper link: https://arxiv.org/abs/2411.10504

YouTube Result: I don't know.

Paper: SLAM3R: Real-Time Dense Scene Reconstruction from Monocular RGB Videos

Paper link: https://arxiv.org/abs/2412.09401

YouTube Result: I don't know.

Paper: MASt3R-SLAM: Real-Time Dense SLAM with 3D Reconstruction Priors

Paper link: https://arxiv.org/abs/2412.12392

YouTube Result: 1. **Motivation**: The study aims to develop a real-time system for dense SLAM

(Simultaneous Localization and Mapping) using a single camera, addressing the need for robust and efficient 3D mapping techniques in real-world scenarios.

2. **Novelty**: The novelty of the study lies in the introduction of a powerful 3D reconstruction prior called MASt3R, which enhances the performance of SLAM systems by providing robust results from real-world video inputs.

3. **Main Findings**: The findings indicate that the MASt3R-SLAM system is capable of generating dense 3D reconstructions in real-time, showcasing its efficacy through visual examples of 3D point maps and camera pose tracking in practical applications.

4. **Video Title**: MASt3R-SLAM: Real-Time Dense 3D Mapping with Priors

5. **Video Link**: https://www.youtube.com/watch?v=fSu0X9xsOqY

Paper: Self-Supervised Cross-View Correspondence with Predictive Cycle Consistency

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: BADGR: Bundle Adjustment Diffusion Conditioned by Gradients for Wide-Baseline Floor Plan Reconstruction

Paper link: https://arxiv.org/abs/2503.19340

YouTube Result: I don't know.

Paper: Light3R-SfM: Towards Feed-forward Structure-from-Motion

Paper link: https://arxiv.org/abs/2501.14914

YouTube Result: I don't know.

Paper: Full-DoF Egomotion Estimation for Event Cameras Using Geometric Solvers

Paper link: https://arxiv.org/abs/2503.03307

YouTube Result: I don't know.

Paper: Active Hyperspectral Imaging Using an Event Camera

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Shape Abstraction via Marching Differentiable Support Functions

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Implicit Correspondence Learning for Image-to-Point Cloud Registration

Paper link: https://arxiv.org/abs/2307.07693

YouTube Result: I don't know.

Paper: Glossy Object Reconstruction with Cost-effective Polarized Acquisition

Paper link: https://arxiv.org/abs/2504.07025

YouTube Result: I don't know.

Paper: Universal Scene Graph Generation

Paper link: https://arxiv.org/abs/2504.01924

YouTube Result: I don't know.

Paper: ICP: Immediate Compensation Pruning for Mid-to-high Sparsity

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Can Generative Video Models Help Pose Estimation?

Paper link: https://arxiv.org/abs/2412.16155

YouTube Result: 1. **Motivation**: The motivation behind the study is to explore the capabilities of generative models in enhancing human pose estimation, which is a crucial aspect of computer vision and has applications in various fields, including robotics, animation, and healthcare.

2. **Novelty**: The novel aspect of the study lies in the integration of generative models and physics simulation to improve the accuracy and efficiency of human pose estimation techniques. This approach offers a new way to augment data and refine pose estimation methods beyond traditional techniques.

3. **Main Findings**: The study demonstrates that generative models can significantly enhance the quality of human pose estimation by providing richer and more diverse training data, thus potentially leading to better performance in real-world applications.

4. **Video Title**: Improving Human Pose Estimation with Generative Models and Physics Simulation

5. **Video Link**: [Improving Human Pose Estimation with Generative Models and Physics Simulation](https://www.youtube.com/watch?v=Aqv_v8yis1I)

Paper: Enduring, Efficient and Robust Trajectory Prediction Attack in Autonomous Driving via Optimization-Driven Multi-Frame Perturbation Framework

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: GEN3C: 3D-Informed World-Consistent Video Generation with Precise Camera Control

Paper link: https://arxiv.org/abs/2503.03751

YouTube Result: 1. **Motivation**: The study behind GEN3C is motivated by the challenges faced in

AI video generation, particularly issues like objects disappearing or appearing unexpectedly and the difficulties in achieving smooth camera movements.

2. **Novelty**: The novel aspect of the GEN3C model is its use of a 3D cache, which acts as a memory bank of point clouds derived from depth predictions of scene images or past frames. This allows for improved consistency and camera control in video generation.

3. **Main Findings**: GEN3C significantly enhances video generation by maintaining 3D consistency and allowing for precise camera control, thereby avoiding the common pitfalls of traditional AI video tools.

4. **Video Title**: GEN3C: 3D-Informed World-Consistent Video Generation with Precise Camera Control (Paper Walkthrough)

5. **Video Link**: [Watch here](https://www.youtube.com/watch?v=Q80Mgm-0JCM)

Paper: SpatialLLM: A Compound 3D-Informed Design towards Spatially-Intelligent Large Multimodal Models

Paper link: https://arxiv.org/abs/2505.00788

YouTube Result: I don't know.

Paper: DriveGPT4-V2: Harnessing Large Language Model Capabilities for Enhanced Closed-Loop Autonomous Driving

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: OmniManip: Towards General Robotic Manipulation via Object-Centric Interaction Primitives as Spatial Constraints

Paper link: https://arxiv.org/abs/2501.03841

YouTube Result: 1. **Motivation**: The study aims to address the challenge of teaching robots to perform precise manipulation tasks in messy and unstructured environments, where traditional methods struggle due to the variability in object arrangements and task requirements.

2. **Novelty**: The novel aspect of the study is the introduction of an object-centric representation, which focuses on how objects function in specific tasks rather than treating them as static entities. This approach emphasizes spatial relationships and interactions, teaching robots to understand objects in the context of their use.

3. **Main Findings**: The findings highlight that by using vision language models (VMS) combined with 3D spatial understanding, the OmniManip system can perform tasks such as pouring tea or inserting a pen into a holder with high accuracy. The system effectively bridges the gap between high-level reasoning and low-level precision needed for complex manipulation tasks.

4. **Video Title**: OmniManip: Robotic Manipulation via Object-Centric Interaction Primitives as Spatial Constraints

5. **Video Link**: [Watch the video](https://www.youtube.com/watch?v=6786hqH894E)

Paper: Generating 6DoF Object Manipulation Trajectories from Action Description in Egocentric Vision

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Erase Diffusion: Empowering Object Removal Through Calibrating Diffusion Pathways

Paper link: https://arxiv.org/abs/2503.07026

YouTube Result: I don't know.

Paper: GigaHands: A Massive Annotated Dataset of Bimanual Hand Activities

Paper link: https://arxiv.org/abs/2412.04244

YouTube Result: 1. **Motivation**: The study aims to address the significant challenge of understanding and replicating human dexterity in AI and robotics, particularly focusing on the intricate manipulation of objects using both hands. A major bottleneck identified is the lack of suitable, large-scale, and accurately annotated datasets for bimanual activities.

2. **Novelty**: The novelty of the study lies in the creation of "Giga Hands," which is a massive annotated dataset specifically designed to capture the complexity of bimanual hand-object interactions. This dataset overcomes the limitations of existing datasets by providing a more diverse and accurate representation of 3D hand activities.

3. **Main Findings**: The study highlights the inadequacies of existing datasets, including issues with 3D accuracy, realism, and annotations. It stresses that many current datasets fail to provide reliable 3D information and lack detailed descriptions necessary for training models effectively.

4. **Video Title**: Seminar: Advancing Large-Scale Dataset for Bimanual Hand-Object Interaction

5. **Video Link**: [Watch the video here](https://www.youtube.com/watch?v=qEi9wJ1NKqw)

Paper: FRAME: Floor-aligned Representation for Avatar Motion from Egocentric Video

Paper link: https://arxiv.org/abs/2503.23094

YouTube Result: I don't know.

Paper: Lifting Motion to the 3D World via 2D Diffusion

Paper link: https://arxiv.org/abs/2411.18808

YouTube Result: 1. **Motivation**: The study aims to generate 3D motion from 2D pose sequences without the need for training on any 3D motion data, addressing the limited availability of datasets containing consistent multi-view 3D motion data.

2. **Novelty**: The novel aspect of the study is the introduction of the MV lift framework, which reformulates 3D motion estimation as generating consistent multi-view 2D pose sequences using only readily available single-view 2D sequences.

3. **Main Findings**: The approach successfully estimates complete 3D motion, including joint rotations and root translations, through a four-stage process that progressively establishes multi-view consistency while utilizing 2D motion diffusion techniques.

4. **Video Title**: Lifting Motion to the 3D World via 2D Diffusion

5. **Video Link**: [Lifting Motion to the 3D World via 2D Diffusion](https://www.youtube.com/watch?v=nffTJHUR8yw)

Paper: RGBAvatar: Reduced Gaussian Blendshapes for Online Modeling of Head Avatars

Paper link: https://arxiv.org/abs/2503.12886

YouTube Result: 1. **Motivation**: The study aims to improve the efficiency and detail of head avatar animation by developing a method that can quickly reconstruct a head avatar model from a short monocular video. This is particularly useful for real-time applications where immediate visual feedback is necessary.

2. **Novelty**: The novel aspect of the study is the ability to reconstruct a detailed head avatar model in approximately 80 seconds from a 2 to 3 minute video, achieving real-time animation at around 400fps. Additionally, it captures finer details such as teeth, eyeglass reflections, and deeper wrinkles compared to existing Gaussian-based methods.

3. **Main Findings**: The method shows significant improvements in expressive facial animations with fewer blend shapes (20 blend shapes are sufficient to capture essential details like wrinkles). It also allows for on-the-fly reconstruction, which enables immediate use for self-reenactment without post-processing.

4. **Video Title**: RGBAvatar Reduced Gaussian Blendshapes

5. **Video Link**: [Watch the video here](https://www.youtube.com/watch?v=r0Rl6t-Btlc)

Paper: EnergyMoGen: Compositional Human Motion Generation with Energy-Based Diffusion Model in Latent Space

Paper link: https://arxiv.org/abs/2412.14706

YouTube Result: I don't know.

Paper: FIction: 4D Future Interaction Prediction from Video

Paper link: https://arxiv.org/abs/2412.00932

YouTube Result: I don't know.

Paper: FoundHand: Large-Scale Domain-Specific Learning for Controllable Hand Image Generation

Paper link: https://arxiv.org/abs/2412.02690

YouTube Result: I don't know.

Paper: 4Real-Video: Learning Generalizable Photo-Realistic 4D Video Diffusion

Paper link: https://arxiv.org/abs/2412.04462

YouTube Result: I don't know.

Paper: EvEnhancer: Empowering Effectiveness, Efficiency and Generalizability for Continuous Space-Time Video Super-Resolution with Events

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Mamba as a Bridge: Where Vision Foundation Models Meet Vision Language Models for Domain-Generalized Semantic Segmentation

Paper link: https://arxiv.org/abs/2504.03193

YouTube Result: I don't know.

Paper: Balanced Rate-Distortion Optimization in Learned Image Compression

Paper link: https://arxiv.org/abs/2502.20161

YouTube Result: I don't know.

Paper: DyFo: A Training-Free Dynamic Focus Visual Search for Enhancing LMMs in Fine-Grained Visual Understanding

Paper link: https://arxiv.org/abs/2504.14920

YouTube Result: I don't know.

Paper: LP-Diff: Towards Improved Restoration of Real-World Degraded License Plate

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: CoSER: Towards Consistent Dense Multiview Text-to-Image Generator for 3D Creation

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: CoMM: A Coherent Interleaved Image-Text Dataset for Multimodal Understanding and Generation

Paper link: https://arxiv.org/abs/2406.10462

YouTube Result: I don't know.

Paper: UniRestore: Unified Perceptual and Task-Oriented Image Restoration Model Using Diffusion

Prior

Paper link: https://arxiv.org/abs/2501.13134

YouTube Result: I don't know.

Paper: Optimizing for the Shortest Path in Denoising Diffusion Model

Paper link: https://arxiv.org/abs/2503.03265

YouTube Result: 1. **Motivation**: The study aims to optimize the image corruption procedure in diffusion models by determining the shortest path in the space of probability distributions, which is crucial for improving image generation quality.

2. **Novelty**: The novel aspect of the study is the introduction of the shortest path diffusion (SPD) framework, which uses the Fisher metric to analytically compute the shortest path between Gaussian distributions and proposes an approximation for non-Gaussian cases.

3. **Main Findings**: The research finds that the optimal corruption procedure corresponds to the shortest path, which significantly enhances the performance of diffusion models, as it outperforms previous methods based on image blurring without requiring hyperparameter tuning. The SPD shows improvements on datasets such as CIFAR-10 and ImageNet 64x64 and introduces various enhancements in diffusion models.

4. **Video Title**: Image generation with shortest path diffusion - ArXiv:2306.00501

5. **Video Link**: [Watch here](https://www.youtube.com/watch?v=Ib0QflUDPUM)

Paper: TinyFusion: Diffusion Transformers Learned Shallow

Paper link: https://arxiv.org/abs/2412.01199

YouTube Result: I don't know.

Paper: Style-Editor: Text-driven Object-centric Style Editing

Paper link: https://arxiv.org/abs/2408.08461

YouTube Result: I don't know.

Paper: MUSt3R: Multi-view Network for Stereo 3D Reconstruction

Paper link: https://arxiv.org/abs/2503.01661

YouTube Result: I don't know.

Paper: DPU: Dynamic Prototype Updating for Multimodal Out-of-Distribution Detection

Paper link: https://arxiv.org/abs/2411.08227

YouTube Result: I don't know.

Paper: Focus-N-Fix: Region-Aware Fine-Tuning for Text-to-Image Generation

Paper link: https://arxiv.org/abs/2501.06481

YouTube Result: I don't know.

Paper: DefectFill: Realistic Defect Generation with Inpainting Diffusion Model for Visual Inspection

Paper link: https://arxiv.org/abs/2503.13985

YouTube Result: I don't know.

Paper: Circumventing Shortcuts in Audio-visual Deepfake Detection Datasets with Unsupervised Learning

Paper link: https://arxiv.org/abs/2412.00175

YouTube Result: I don't know.

Paper: UIBDiffusion: Universal Imperceptible Backdoor Attack for Diffusion Models

Paper link: https://arxiv.org/abs/2412.11441

YouTube Result: I don't know.

Paper: Open-Vocabulary Functional 3D Scene Graphs for Real-World Indoor Spaces

Paper link: https://arxiv.org/abs/2503.19199

YouTube Result: 1. **Motivation**: The study aims to address the limitations of current mapping techniques, such as V maps and concept fusion, which struggle with scalability in larger scenes due to high storage overhead. The motivation is to improve the representation and navigation in large-scale environments through enhanced 3D scene graphs.

2. **Novelty**: The novel aspect of the study is the introduction of Hierarchical Open-Vocabulary 3D Scene Graphs (HOV-SG), which enrich traditional 3D scene graphs with open vocabulary features and explicitly model hierarchical semantics, thereby allowing for better mapping in large-scale environments.

3. **Main Findings**: The study presents a two-stage pipeline for creating segment-level open vocabulary maps and constructing a scene graph from RGBD sequences. It demonstrates how to efficiently merge per-frame segments based on global features and apply clustering techniques to build a more accurate representation of complex indoor environments.

4. **Video Title**: HOV-SG: Hierarchical Open-Vocabulary 3D Scene Graphs for Language-Grounded Robot Navigation (RSS'24)

5. **Video Link**: [HOV-SG Video](https://www.youtube.com/watch?v=GC-Q0ekO9qg)

Paper: AnySat: One Earth Observation Model for Many Resolutions, Scales, and Modalities

Paper link: https://arxiv.org/abs/2412.14123

YouTube Result: I don't know.

Paper: All Languages Matter: Evaluating LMMs on Culturally Diverse 100 Languages

Paper link: https://arxiv.org/abs/2411.16508

YouTube Result: I don't know.

Paper: CL-MoE: Enhancing Multimodal Large Language Model with Dual Momentum Mixture-of-Experts for Continual Visual Question Answering

Paper link: https://arxiv.org/abs/2503.00413

YouTube Result: I don't know.

Paper: Improving Personalized Search with Regularized Low-Rank Parameter Updates

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: PhD: A ChatGPT-Prompted Visual Hallucination Evaluation Dataset

Paper link: https://arxiv.org/abs/2403.11116

YouTube Result: I don't know.

Paper: O-TPT: Orthogonality Constraints for Calibrating Test-time Prompt Tuning in Vision-Language Models

Paper link: https://arxiv.org/abs/2503.12096

YouTube Result: I don't know.

Paper: RLAIF-V: Open-Source AI Feedback Leads to Super GPT-4V Trustworthiness

Paper link: https://arxiv.org/abs/2405.17220

YouTube Result: I don't know.

Paper: F^3OCUS - Federated Finetuning of Vision-Language Foundation Models with Optimal Client Layer Updating Strategy via Multi-objective Meta-Heuristics

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: CARE Transformer: Mobile-Friendly Linear Visual Transformer via Decoupled Dual Interaction

Paper link: https://arxiv.org/abs/2411.16170

YouTube Result: I don't know.

Paper: Coeff-Tuning: A Graph Filter Subspace View for Tuning Attention-Based Large Models

Paper link: https://arxiv.org/abs/2503.18337

YouTube Result: I don't know.

Paper: Generative Modeling of Class Probability for Multi-Modal Representation Learning

Paper link: https://arxiv.org/abs/2503.17417

YouTube Result: I don't know.

Paper: STING-BEE: Towards Vision-Language Model for Real-World X-ray Baggage Security Inspection

Paper link: https://arxiv.org/abs/2504.02823

YouTube Result: I don't know.

Paper: Perceptually Accurate 3D Talking Head Generation: New Definitions, Speech-Mesh Representation, and Evaluation Metrics

Paper link: https://arxiv.org/abs/2503.20308

YouTube Result: I don't know.

Paper: PGC: Physics-Based Gaussian Cloth from a Single Pose

Paper link: https://arxiv.org/abs/2503.20779

YouTube Result: 1. **Motivation**: The study aims to reconstruct simulation-ready garments with intricate appearance from a single pose, addressing the challenges of garment representation in 3D modeling and rendering.

2. **Novelty**: The approach utilizes a hybrid mesh embedded with 3D Gaussian splats, allowing for pose generalization and integrating physics-based simulation and rendering techniques while capturing fine garment details.

3. **Main Findings**: The proposed method outperforms traditional techniques by combining the advantages of physics-based rendering with detailed shading, resulting in realistic garment appearances even when deformed into novel poses. The study demonstrates superior performance in rendering compared to existing methods.

4. **Video Title**: PGC: Physics-based Gaussian Cloth from a Single Pose

5. **Video Link**: [Watch here](https://www.youtube.com/watch?v=Pi4Kw2wUBSU)

Paper: Is this Generated Person Existed in Real-world? Fine-grained Detecting and Calibrating Abnormal Human-body

Paper link: https://arxiv.org/abs/2411.14205

YouTube Result: I don't know.

Paper: ARM: Appearance Reconstruction Model for Relightable 3D Generation

Paper link: https://arxiv.org/abs/2411.10825

YouTube Result: I don't know.

Paper: CADDreamer: CAD Object Generation from Single-view Images

Paper link: https://arxiv.org/abs/2502.20732

YouTube Result: I don't know.

Paper: High-fidelity 3D Object Generation from Single Image with RGBN-Volume Gaussian Reconstruction Model

Paper link: https://arxiv.org/abs/2504.01512

YouTube Result: I don't know.

Paper: Panorama Generation From NFoV Image Done Right

Paper link: https://arxiv.org/abs/2503.18420

YouTube Result: I don't know.

Paper: World-consistent Video Diffusion with Explicit 3D Modeling

Paper link: https://arxiv.org/abs/2503.07135

YouTube Result: I don't know.

Paper: Improving Gaussian Splatting with Localized Points Management

Paper link: https://arxiv.org/abs/2411.08373

YouTube Result: I don't know.

Paper: Event Ellipsometer: Event-based Mueller-Matrix Video Imaging

Paper link: https://arxiv.org/abs/2411.17313

YouTube Result: I don't know.

Paper: All-directional Disparity Estimation for Real-world QPD Images

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Reconstructing People, Places, and Cameras

Paper link: https://arxiv.org/abs/2412.17806

YouTube Result: I don't know.

Paper: SeCap: Self-Calibrating and Adaptive Prompts for Cross-view Person Re-Identification in Aerial-Ground Networks

Paper link: https://arxiv.org/abs/2503.06965

YouTube Result: I don't know.

Paper: Sonata: Self-Supervised Learning of Reliable Point Representations

Paper link: https://arxiv.org/abs/2503.16429

YouTube Result: I don't know.

Paper: BWFormer: Building Wireframe Reconstruction from Airborne LiDAR Point Cloud with Transformer

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: DexGrasp Anything: Towards Universal Robotic Dexterous Grasping with Physics Awareness

Paper link: https://arxiv.org/abs/2503.08257

YouTube Result: 1. **Motivation**: The study aims to develop a dextrous robotic hand capable of grasping any object, which is essential for creating general-purpose embodied robots. The challenge lies in the complexity of hand degrees of freedom and the diversity of objects.

2. **Novelty**: The novel aspect of the study is the integration of physical constraints into both the training and sampling phases of a diffusion-based generative model, which enhances the process of generating robust and high-quality grasp poses.

3. **Main Findings**: The method, DexGrasp Anything, successfully combines object point cloud and shadow hand pose parameters to guide the generation of suitable hand poses for grasping. By utilizing physical constraints during both training and sampling, the approach enables effective grasping of a wide variety of objects.

4. **Video Title**: DexGrasp Anything: Towards Universal Robotic Dexterous Grasping with Physics Awareness (CVPR 2025)

5. **Video Link**: [Watch here](https://www.youtube.com/watch?v=KNEY_LKG5Y4)

Paper: H-MoRe: Learning Human-centric Motion Representation for Action Analysis

Paper link: https://arxiv.org/abs/2504.10676

YouTube Result: I don't know.

Paper: Tracktention: Leveraging Point Tracking to Attend Videos Faster and Better

Paper link: https://arxiv.org/abs/2503.19904

YouTube Result: I don't know.

Paper: Align3R: Aligned Monocular Depth Estimation for Dynamic Videos

Paper link: https://arxiv.org/abs/2412.03079

YouTube Result: I don't know.

Paper: Meta-Learning Hyperparameters for Parameter Efficient Fine-Tuning

Paper link: https://arxiv.org/abs/2008.05984

YouTube Result: 1. **Motivation**: The study addresses the challenge of fine-tuning natural language models in a parameter-efficient manner while maintaining accuracy. It emphasizes the need for few-shot learning, a capability that allows models to generalize from a limited number of examples, which is particularly useful for users with constrained resources and limited labeled data.

2. **Novelty**: The study explores the application of meta-learning to enhance few-shot learning capabilities in natural language processing (NLP) models. This approach aims to optimize the fine-tuning process by leveraging meta-learning techniques to improve performance with minimal data.

3. **Main Findings**: The findings suggest that using meta-learning can significantly enhance the ability of models to adapt to new tasks with limited examples. This is particularly beneficial for cloud-based machine learning services that cater to diverse user needs, enabling the provision of highly accurate models that are adaptable for various tasks.

4. **Video Title**: [AutoMLConf'22]: Meta-Adapters: Parameter Efficient Few-shot Fine-tuning through Meta-Learning

5. **Video Link**: [Watch the Video](https://www.youtube.com/watch?v=PdB80toLiAw)

Paper: Understanding Multi-layered Transmission Matrices

Paper link: https://arxiv.org/abs/2410.23864

YouTube Result: 1. **Motivation**: The motivation behind the study is to understand how wave optics can be treated using a transfer matrix approach, similar to prior discussions on ray optics, in order to analyze multilayered photonic systems more effectively.

2. **Novelty**: The novel aspect of this study is the introduction of wave optics transfer matrix methods, which allows for the systematic calculation of light behavior at multiple interfaces within layered media, incorporating phase changes during propagation.

3. **Main Findings**: The main findings indicate that by applying the transfer matrix approach, one can determine the effects of multiple layers on light transmission and reflection, using established equations like Fresnel's equations for incident waves and considering phase shifts as waves pass through different media.

4. **Video Title**: nanoHUB-U Nanophotonic Modeling L2.2: Multilayered Photonic Systems: Wave Optics Transfer Matrices

5. **Video Link**: [Watch here](https://www.youtube.com/watch?v=9LibD5qJ_Rs)

Paper: Good, Cheap, and Fast: Overfitted Image Compression with Wasserstein Distortion

Paper link: https://arxiv.org/abs/2412.00505

YouTube Result: I don't know.

Paper: Visual Representation Learning through Causal Intervention for Controllable Image Editing

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Boost Your Human Image Generation Model via Direct Preference Optimization

Paper link: https://arxiv.org/abs/2405.20216

YouTube Result: I don't know.

Paper: ReNeg: Learning Negative Embedding with Reward Guidance

Paper link: https://arxiv.org/abs/2412.19637

YouTube Result: I don't know.

Paper: STEREO: A Two-Stage Framework for Adversarially Robust Concept Erasing from Text-to-Image Diffusion Models

Paper link: https://arxiv.org/abs/2408.16807

YouTube Result: I don't know.

Paper: Supervising Sound Localization by In-the-wild Egomotion

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Cross-modal Causal Relation Alignment for Video Question Grounding

Paper link: https://arxiv.org/abs/2503.07635

YouTube Result: I don't know.

Paper: Video-MME: The First-Ever Comprehensive Evaluation Benchmark of Multi-modal LLMs in Video Analysis

Paper link: https://arxiv.org/abs/2405.21075

YouTube Result: I don't know.

Paper: Just Dance with pi! A Poly-modal Inductor for Weakly-supervised Video Anomaly Detection

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Learning 4D Panoptic Scene Graph Generation from Rich 2D Visual Scene

Paper link: https://arxiv.org/abs/2503.15019

YouTube Result: I don't know.

Paper: The Scene Language: Representing Scenes with Programs, Words, and Embeddings

Paper link: https://arxiv.org/abs/2410.16770

YouTube Result: I don't know.

Paper: VL-RewardBench: A Challenging Benchmark for Vision-Language Generative Reward Models

Paper link: https://arxiv.org/abs/2503.06260

YouTube Result: I don't know.

Paper: Spatial457: A Diagnostic Benchmark for 6D Spatial Reasoning of Large Mutimodal Models

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Not Only Text: Exploring Compositionality of Visual Representations in Vision-Language Models

Paper link: https://arxiv.org/abs/2503.17142

YouTube Result: I don't know.

Paper: Realistic Test-Time Adaptation of Vision-Language Models

Paper link: https://arxiv.org/abs/2501.03729

YouTube Result: I don't know.

Paper: Comprehensive Information Bottleneck for Unveiling Universal Attribution to Interpret Vision Transformers

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: T2ICount: Enhancing Cross-modal Understanding for Zero-Shot Counting

Paper link: https://arxiv.org/abs/2502.20625

YouTube Result: I don't know.

Paper: WISH: Weakly Supervised Instance Segmentation using Heterogeneous Labels

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Dataset Distillation with Neural Characteristic Function: A Minmax Perspective

Paper link: https://arxiv.org/abs/2502.20653

YouTube Result: I don't know.

Paper: SoMA: Singular Value Decomposed Minor Components Adaptation for Domain Generalizable Representation Learning

Paper link: https://arxiv.org/abs/2412.04077

YouTube Result: I don't know.

Paper: Free-viewpoint Human Animation with Pose-correlated Reference Selection

Paper link: https://arxiv.org/abs/2412.17290

YouTube Result: I don't know.

Paper: Real-time High-fidelity Gaussian Human Avatars with Position-based Interpolation of Spatially Distributed MLPs

Paper link: https://arxiv.org/abs/2504.12909

YouTube Result: I don't know.

Paper: Material Anything: Generating Materials for Any 3D Object via Diffusion

Paper link: https://arxiv.org/abs/2411.15138

YouTube Result: I don't know.

Paper: Generative Densification: Learning to Densify Gaussians for High-Fidelity Generalizable 3D Reconstruction

Paper link: https://arxiv.org/abs/2412.06234

YouTube Result: I don't know.

Paper: Event Fields: Capturing Light Fields at High Speed, Resolution, and Dynamic Range

Paper link: https://arxiv.org/abs/2412.06191

YouTube Result: I don't know.

Paper: IncEventGS: Pose-Free Gaussian Splatting from a Single Event Camera

Paper link: https://arxiv.org/abs/2410.08107

YouTube Result: I don't know.

Paper: HELVIPAD: A Real-World Dataset for Omnidirectional Stereo Depth Estimation

Paper link: https://arxiv.org/abs/2503.23502

YouTube Result: I don't know.

Paper: Order-One Rolling Shutter Cameras

Paper link: https://arxiv.org/abs/2403.11295

YouTube Result: I don't know.

Paper: Simulator HC: Regression-based Online Simulation of Starting Problem-Solution Pairs for Homotopy Continuation in Geometric Vision

Paper link: https://arxiv.org/abs/2411.03745

YouTube Result: I don't know.

Paper: GaussianUDF: Inferring Unsigned Distance Functions through 3D Gaussian Splatting

Paper link: https://arxiv.org/abs/2503.19458

YouTube Result: I don't know.

Paper: Doppelgangers++: Improved Visual Disambiguation with Geometric 3D Features

Paper link: https://arxiv.org/abs/2412.05826

YouTube Result: I don't know.

Paper: MITracker: Multi-View Integration for Visual Object Tracking

Paper link: https://arxiv.org/abs/2502.20111

YouTube Result: I don't know.

Paper: Ev-3DOD: Pushing the Temporal Boundaries of 3D Object Detection with Event Cameras

Paper link: https://arxiv.org/abs/2502.19630

YouTube Result: I don't know.

Paper: Deep Change Monitoring: A Hyperbolic Representative Learning Framework and a Dataset for Long-term Fine-grained Tree Change Detection

Paper link: https://arxiv.org/abs/2503.00643

YouTube Result: I don't know.

Paper: SplatFlow: Self-Supervised Dynamic Gaussian Splatting in Neural Motion Flow Field for Autonomous Driving

Paper link: https://arxiv.org/abs/2411.15482

YouTube Result: I don't know.

Paper: Towards Autonomous Micromobility through Scalable Urban Simulation

Paper link: https://arxiv.org/abs/2505.00690

YouTube Result: I don't know.

Paper: RoboTwin: Dual-Arm Robot Benchmark with Generative Digital Twins

Paper link: https://arxiv.org/abs/2504.13059

YouTube Result: I don't know.

Paper: End-to-End HOI Reconstruction Transformer with Graph-based Encoding

Paper link: https://arxiv.org/abs/2503.06012

YouTube Result: I don't know.

Paper: Dyn-HaMR: Recovering 4D Interacting Hand Motion from a Dynamic Camera

Paper link: https://arxiv.org/abs/2412.12861

YouTube Result: I don't know.

Paper: MotionPRO: Exploring the Role of Pressure in Human MoCap and Beyond

Paper link: https://arxiv.org/abs/2504.05046

YouTube Result: I don't know.

Paper: UniPose: A Unified Multimodal Framework for Human Pose Comprehension, Generation and Editing

Paper link: https://arxiv.org/abs/2411.16781

YouTube Result: I don't know.

Paper: Unified Reconstruction of Static and Dynamic Scenes from Events

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: FreePCA: Integrating Consistency Information across Long-short Frames in Training-free Long Video Generation via Principal Component Analysis

Paper link: https://arxiv.org/abs/2505.01172

YouTube Result: I don't know.

Paper: All-Optical Nonlinear Diffractive Deep Network for Ultrafast Image Denoising

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: FlexiDiT: Your Diffusion Transformer Can Easily Generate High-Quality Samples with Less Compute

Paper link: https://arxiv.org/abs/2502.20126

YouTube Result: I don't know.

Paper: Gaze-LLE: Gaze Target Estimation via Large-Scale Learned Encoders

Paper link: https://arxiv.org/abs/2412.09586

YouTube Result: I don't know.

Paper: Which Viewpoint Shows it Best? Language for Weakly Supervising View Selection in Multi-view Instructional Videos

Paper link: https://arxiv.org/abs/2411.08753

YouTube Result: I don't know.

Paper: CASAGPT: Cuboid Arrangement and Scene Assembly for Interior Design

Paper link: https://arxiv.org/abs/2504.19478

YouTube Result: I don't know.

Paper: Revealing Key Details to See Differences: A Novel Prototypical Perspective for Skeleton-based Action Recognition

Paper link: https://arxiv.org/abs/2411.18941

YouTube Result: I don't know.

Paper: Octopus: Alleviating Hallucination via Dynamic Contrastive Decoding

Paper link: https://arxiv.org/abs/2503.00361

YouTube Result: I don't know.

Paper: TIDE: Training Locally Interpretable Domain Generalization Models Enables Test-time Correction

Paper link: https://arxiv.org/abs/2411.16788

YouTube Result: I don't know.

Paper: UCOD-DPL: Unsupervised Camouflaged Object Detection via Dynamic Pseudo-label Learning

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: SURGEON: Memory-Adaptive Fully Test-Time Adaptation via Dynamic Activation Sparsity

Paper link: https://arxiv.org/abs/2503.20354

YouTube Result: I don't know.

Paper: ROLL: Robust Noisy Pseudo-label Learning for Multi-View Clustering with Noisy Correspondence

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Do Computer Vision Foundation Models Learn the Low-level Characteristics of the Human Visual System?

Paper link: https://arxiv.org/abs/2502.20256

YouTube Result: I don't know.

Paper: Distraction is All You Need for Multimodal Large Language Model Jailbreaking

Paper link: https://arxiv.org/abs/2502.10794

YouTube Result: I don't know.

Paper: Context-Aware Multimodal Pretraining

Paper link: https://arxiv.org/abs/2505.03315

YouTube Result: I don't know.

Paper: LaTexBlend: Scaling Multi-concept Customized Generation with Latent Textual Blending

Paper link: https://arxiv.org/abs/2503.06956

YouTube Result: I don't know.

Paper: How Do I Do That? Synthesizing 3D Hand Motion and Contacts for Everyday Interactions

Paper link: https://arxiv.org/abs/2504.12284

YouTube Result: I don't know.

Paper: VILA-M3: Enhancing Vision-Language Models with Medical Expert Knowledge

Paper link: https://arxiv.org/abs/2411.12915

YouTube Result: I don't know.

Paper: DiffCAM: Data-Driven Saliency Maps by Capturing Feature Differences

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Text-guided Sparse Voxel Pruning for Efficient 3D Visual Grounding

Paper link: https://arxiv.org/abs/2502.10392

YouTube Result: I don't know.

Paper: NLPrompt: Noise-Label Prompt Learning for Vision-Language Models

Paper link: https://arxiv.org/abs/2412.01256

YouTube Result: I don't know.

Paper: Understanding Multi-Task Activities from Single-Task Videos

Paper link: https://arxiv.org/abs/2503.18223

YouTube Result: I don't know.

Paper: Few-shot Implicit Function Generation via Equivariance

Paper link: https://arxiv.org/abs/2501.01601

YouTube Result: I don't know.

Paper: Instant Gaussian Stream: Fast and Generalizable Streaming of Dynamic Scene Reconstruction via Gaussian Splatting

Paper link: https://arxiv.org/abs/2503.16979

YouTube Result: I don't know.

Paper: InPO: Inversion Preference Optimization with Reparametrized DDIM for Efficient Diffusion Model Alignment

Paper link: https://arxiv.org/abs/2503.18454

YouTube Result: I don't know.

Paper: Galaxy Walker: Geometry-aware VLMs For Galaxy-scale Understanding

Paper link: https://arxiv.org/abs/2503.18578

YouTube Result: I don't know.

Paper: StyleSSP: Sampling StartPoint Enhancement for Training-free Diffusion-based Method for Style Transfer

Paper link: https://arxiv.org/abs/2501.11319

YouTube Result: I don't know.

Paper: Detecting Backdoor Attacks in Federated Learning via Direction Alignment Inspection

Paper link: https://arxiv.org/abs/2503.07978

YouTube Result: I don't know.

Paper: Empowering Vector Graphics with Consistently Arbitrary Viewing and View-dependent Visibility

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: LeviTor: 3D Trajectory Oriented Image-to-Video Synthesis

Paper link: https://arxiv.org/abs/2412.15214

YouTube Result: I don't know.

Paper: SSHNet: Unsupervised Cross-modal Homography Estimation via Problem Reformulation and Split Optimization

Paper link: https://arxiv.org/abs/2409.17993

YouTube Result: I don't know.

Paper: Video Depth Anything: Consistent Depth Estimation for Super-Long Videos

Paper link: https://arxiv.org/abs/2501.12375

YouTube Result: I don't know.

Paper: GenVDM: Generating Vector Displacement Maps From a Single Image

Paper link: https://arxiv.org/abs/2503.00605

YouTube Result: I don't know.

Paper: Cubify Anything: Scaling Indoor 3D Object Detection

Paper link: https://arxiv.org/abs/2412.04458

YouTube Result: I don't know.

Paper: 3D Convex Splatting: Radiance Field Rendering with 3D Smooth Convexes

Paper link: https://arxiv.org/abs/2411.14974

YouTube Result: I don't know.

Paper: MetricGrids: Arbitrary Nonlinear Approximation with Elementary Metric Grids based Implicit Neural Representation

Paper link: https://arxiv.org/abs/2503.10000

YouTube Result: I don't know.

Paper: Project-Probe-Aggregate: Efficient Fine-Tuning for Group Robustness

Paper link: https://arxiv.org/abs/2503.09487

YouTube Result: I don't know.

Paper: NexusGS: Sparse View Synthesis with Epipolar Depth Priors in 3D Gaussian Splatting

Paper link: https://arxiv.org/abs/2503.18794

YouTube Result: I don't know.

Paper: Towards In-the-wild 3D Plane Reconstruction from a Single Image

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: MATCHA: Towards Matching Anything

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: A Unified Approach to Interpreting Self-supervised Pre-training Methods for 3D Point Clouds via Interactions

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: A Polarization-Aided Transformer for Image Deblurring via Motion Vector Decomposition

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: SmartCLIP: Modular Vision-language Alignment with Identification Guarantees

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: ImagineFSL: Self-Supervised Pretraining Matters on Imagined Base Set for VLM-based Few-shot Learning

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Multi-Label Prototype Visual Spatial Search for Weakly Supervised Semantic Segmentation

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: DAMM-Diffusion: Learning Divergence-Aware Multi-Modal Diffusion Model for Nanoparticles Distribution Prediction

Paper link: https://arxiv.org/abs/2503.09491

YouTube Result: I don't know.

Paper: DIV-FF: Dynamic Image-Video Feature Fields For Environment Understanding in Egocentric Videos

Paper link: https://arxiv.org/abs/2503.08344

YouTube Result: I don't know.

Paper: BlenderGym: Benchmarking Foundational Model Systems for Graphics Editing

Paper link: https://arxiv.org/abs/2504.01786

YouTube Result: I don't know.

Paper: Creating Your Editable 3D Photorealistic Avatar with Tetrahedron-constrained Gaussian Splatting

Paper link: https://arxiv.org/abs/2504.20403

YouTube Result: I don't know.

Paper: Efficient Motion-Aware Video MLLM

Paper link: https://arxiv.org/abs/2504.13074

YouTube Result: I don't know.

Paper: Structure-from-Motion with a Non-Parametric Camera Model

Paper link: https://arxiv.org/abs/2309.17054

YouTube Result: I don't know.

Paper: Advancing Multiple Instance Learning with Continual Learning for Whole Slide Imaging

Paper link: https://arxiv.org/abs/2408.15032

YouTube Result: I don't know.

Paper: Polarized Color Screen Matting

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Samba: A Unified Mamba-based Framework for General Salient Object Detection

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: NSD-Imagery: A Benchmark Dataset for Extending fMRI Vision Decoding Methods to Mental Imagery

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Point-to-Region Loss for Semi-Supervised Point-Based Crowd Counting

Paper link: Not found on arXiv

YouTube Result: I don't know.

Paper: Blurred LiDAR for Sharper 3D: Robust Handheld 3D Scanning with Diffuse LiDAR and RGB

Paper link: https://arxiv.org/abs/2411.19474

YouTube Result: I don't know.