Novelty Assessment Report

Paper: Image Quality Assessment for Embodied AI
PDF URL: https://openreview.net/pdf?id=azj53PLJRL
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-27

Abstract

Embodied AI has developed rapidly in recent years, but it is still mainly deployed in laboratories, with various distortions in the Real-world limiting its application. Traditionally, Image Quality Assessment (IQA) methods are applied to predict human preferences for distorted images; however, there is no IQA method to assess the usability of an image in embodied tasks, namely, the perceptual quality for robots. To provide accurate and reliable quality indicators for future embodied scenarios, we first propose the topic: IQA for Embodied AI. Specifically, we (1) based on the Mertonian system and meta-cognitive theory, constructed a perception-cognition-decision-execution pipeline and defined a comprehensive subjective score collection process; (2) established the Embodied-IQA database, containing over 30k reference/distorted image pairs, with more than 5m fine-grained annotations provided by Vision Language Models/Vision Language Action-models/Real-world robots; (3) trained and validated the performance of mainstream IQA methods on Embodied-IQA, demonstrating the need to develop more accurate quality indicators for Embodied AI. We sincerely hope that through evaluation, we can promote the application of Embodied AI under complex distortions in the Real-world.

Core Task Landscape

This paper addresses: **Image Quality Assessment for Embodied AI Tasks**
A total of **50 papers** were analyzed and organized into a taxonomy with **23 categories**.

Taxonomy Overview

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

- **Quality Assessment Frameworks and Benchmarks**
- **Visual Perception and Representation Learning**
- **Embodied Navigation and Spatial Reasoning**
- **Manipulation and Interaction**
- **Scene Reconstruction and 3D Representation**
- **Generative Models and World Simulation**
- **Simulation Platforms and Datasets**
- **System Integration and Applications**

Complete Taxonomy Tree

- Image Quality Assessment for Embodied AI Tasks Survey Taxonomy
- Quality Assessment Frameworks and Benchmarks
  - Embodied-Specific Quality Assessment ★ (4 papers)
  - [0] Image Quality Assessment for Embodied AI (Anon et al., 2026) View paper
  - [1] Perceptual Quality Assessment for Embodied AI (C Li, 2025) View paper
  - [12] Embodied Image Quality Assessment for Robotic Intelligence (Zhang Jianbo, 2024) View paper
  - [49] RGC-VQA: An Exploration Database for Robotic-Generated Video Quality Assessment (Jin Jianing, 2025) View paper
  - World Model and Generative Content Evaluation (3 papers)
  - [2] Ewmbench: Evaluating scene, motion, and semantic quality in embodied world models (Yue Hu, 2025) View paper
  - [19] World-in-World: World Models in a Closed-Loop World (Zhang JiaHan, 2025) View paper
  - [23] Are Video Models Ready as Zero-Shot Reasoners? An Empirical Study with the MME-CoF Benchmark (Guo Ziyu, 2025) View paper
  - General Visual Quality Assessment (1 papers)
  - [28] Perceptual Visual Quality Assessment in Multimedia Communication (Wei Zhou, 2025) View paper
- Visual Perception and Representation Learning
  - Pre-trained Visual Representations and Foundation Models (2 papers)
  - [3] Where are we in the search for an artificial visual cortex for embodied intelligence? (Majumdar, 2023) View paper
  - [6] Elevating Visual Perception in Multimodal LLMs with Visual Embedding Distillation (Jain, 2024) View paper
  - Multimodal Language-Vision Integration (6 papers)
  - [7] Vision Language Action Models in Robotic Manipulation: A Systematic Review (Akram, 2025) View paper
  - [10] Evaluating Uncertainty and Quality of Visual Language Action-enabled Robots (Valle, 2025) View paper
  - [11] Large Language Models for Robotics: Opportunities, Challenges, and Perspectives (Jiaqi Wang, 2024) View paper
  - [27] VidEgoThink: Assessing Egocentric Video Understanding Capabilities for Embodied AI (Cheng, 2024) View paper
  - [40] RynnEC: Bringing MLLMs into Embodied World (Dang, 2025) View paper
  - [45] XGC-AVis: Towards Audio-Visual Content Understanding with a Multi-Agent Collaborative System (Cao Yuqin, 2025) View paper
  - Specialized Perception Modalities (4 papers)
  - [18] Are Multimodal Large Language Models Ready for Omnidirectional Spatial Reasoning? (Zheng Xu, 2025) View paper

- [25] TextToucher: Fine-Grained Text-to-Touch Generation (Fu Hao, 2024) View paper
- [33] Multi-view Enhanced Few-shot Action Recognition for Embodied Visual Perception (Shuai Li, 2025) View paper
- [37] AirCopBench: A Benchmark for Multi-drone Collaborative Embodied Perception and Reasoning (Jirong Zha, 2025) View paper
- Uncertainty-Aware and Robust Perception (3 papers)
- [30] A lightweight and style-robust neural network for autonomous driving in end side devices (Han Sheng, 2023) View paper
- [31] B-AIS: An Automated Process for Black-box Evaluation of Visual Perception in AI-enabled Software against Domain Semantics (Hamed Barzamini, 2022) View paper
- [44] Perception Matters: Enhancing Embodied AI with Uncertainty-Aware Semantic Segmentation (Honerkamp, 2024) View paper

- Embodied Navigation and Spatial Reasoning
  - Object-Goal and Question-Driven Navigation (4 papers)
  - [5] Objectnav revisited: On evaluation of embodied agents navigating to objects (D Batra, 2020) View paper
  - [26] Embodied Question Answering (Das, 2017) View paper
  - [34] A Survey of Object Goal Navigation: Datasets, Metrics and Methods (Dewei Wang, 2023) View paper
  - [47] Skip-SCAR: Hardware-Friendly High-Quality Embodied Visual Navigation (Cao Yu, 2024) View paper
  - Multimodal Navigation Frameworks (1 papers)
  - [4] Multimodal Perception for Indoor Mobile Robotics Navigation and Safe Manipulation (Yinlong Zhang, 2025) View paper
- Manipulation and Interaction
  - Action-Conditioned Prediction and Control (2 papers)
  - [29] A control-centric benchmark for video prediction (Tian, 2023) View paper
  - [43] Image Quality Assessment in Visual Reinforcement Learning for Fast-moving Targets (Sanghyun Ryoo, 2024) View paper
  - Autonomous Driving and End-to-End Control (1 papers)
  - [21] Generic Simulation Framework for Evaluation Process: Applied to AI-powered Visual Perception System in Autonomous Driving (Wei Xu, 2023) View paper
- Scene Reconstruction and 3D Representation
  - 3D Gaussian Splatting and Rendering (2 papers)
  - [20] GSArch: Breaking Memory Barriers in 3D Gaussian Splatting Training via Architectural Support (Houshu He, 2025) View paper
  - [22] No Redundancy, No Stall: Lightweight Streaming 3D Gaussian Splatting for Real-time Rendering (Wei Linye, 2025) View paper
  - Generalizable and Open-World Scene Reconstruction (2 papers)
  - [46] OGGSplat: Open Gaussian Growing for Generalizable Reconstruction with Expanded Field-of-View (Wang Yan-bo, 2025) View paper
  - [48] Wanderland: Geometrically Grounded Simulation for Open-World Embodied AI (Xinhao Liu, 2025) View paper
  - Single-Image to 3D Scene Generation (1 papers)
  - [50] CC-FMO: Camera-Conditioned Zero-Shot Single Image to 3D Scene Generation with Foundation Model Orchestration (Boshi Tang, 2025) View paper
  - Large-Scale 3D Environment Datasets (1 papers)
  - [9] Habitat-matterport 3d dataset (hm3d): 1000 large-scale 3d environments for embodied ai (Ramakrishnan, 2021) View paper
  - City-Scale and Multi-Agent 3D Generation (1 papers)
  - [42] RAISECity: A Multimodal Agent Framework for Reality-Aligned 3D World Generation at City-Scale (Shengyuan Wang, 2025) View paper
- Generative Models and World Simulation
  - Generative World Models and Mental Exploration (2 papers)
  - [13] Affordance-Based Goal Imagination for Embodied AI Agents (Victor Aregbede, 2024) View paper
  - [24] Generative world explorer (Shu, 2024) View paper
  - Generative Physical AI (1 papers)
  - [17] Generative physical ai in vision: A survey (Liu, 2025) View paper
  - Image Restoration and Enhancement for Embodied Perception (2 papers)
  - [16] Multi-Step Guided Diffusion for Image Restoration on Edge Devices: Toward Lightweight Perception in Embodied AI (Chakravarty, 2025) View paper
  - [39] PerTouch: VLM-Driven Agent for Personalized and Semantic Image Retouching (Zewei Chang, 2025) View paper
- Simulation Platforms and Datasets
  - Embodied AI Simulators and Task Benchmarks (1 papers)
  - [14] A survey of embodied ai: From simulators to research tasks (Duan, 2022) View paper
  - Photorealistic Digital Twins (1 papers)
  - [38] TwinOR: Photorealistic Digital Twins of Dynamic Operating Rooms for Embodied AI Research (Han Zhang, 2025) View paper
- System Integration and Applications
  - Multi-Agent and Communication Systems (3 papers)
  - [15] Embodied AI-Enhanced Vehicular Networks: An Integrated Large Language Models and Reinforcement Learning Method (Zhang Rui-chen, 2025) View paper
  - [35] RAI: Flexible Agent Framework for Embodied AI (Majek Maciej, 2025) View paper
  - [41] Embodied Image Compression (C Li, 2025) View paper
  - Domain-Specific Applications (3 papers)
  - [8] Card: Cross-modal agent framework for generative and editable residential design (Pengyu Zeng, 2025) View paper
  - [32] Shadowplay: An Embodied AI Art Installation (Jesse Josua Benjamin, 2024) View paper
  - [36] Optical performance perception technology in the embodied intelligence-based optical astronomical telescopes (Yiming Zhang, 2025) View paper

Narrative

Core task: Image quality assessment for embodied artificial intelligence tasks. The field spans a diverse set of challenges, from developing quality metrics and benchmarks tailored to embodied settings, to building robust visual perception systems, navigation and manipulation capabilities, scene reconstruction methods, generative world models, and comprehensive simulation platforms. At the top level, the taxonomy organizes work into eight major branches: Quality Assessment Frameworks and Benchmarks focuses on metrics and evaluation protocols specific to embodied contexts (e.g., Perceptual Quality Embodied[1], Embodied Image Quality[12]); Visual Perception and Representation Learning addresses how agents encode and interpret visual input (e.g., Artificial Visual Cortex[3], Visual Embedding Distillation[6]); Embodied Navigation and Spatial Reasoning explores goal-driven movement and spatial understanding (e.g.,

Objectnav Revisited[5], Omnidirectional Spatial Reasoning[18]); Manipulation and Interaction examines physical interaction with objects (e.g., PerTouch[39], TextToucher[25]); Scene Reconstruction and 3D Representation deals with building spatial models (e.g., Lightweight Gaussian Splatting[22], OGGSplat[46]); Generative Models and World Simulation investigates predictive and generative approaches (e.g., Generative Physical AI[17], World in World[19]); Simulation Platforms and Datasets provides testbeds and data resources (e.g., Habitat Matterport[9], Ewmbench[2]); and System Integration and Applications brings these components together in real-world deployments (e.g., Multimodal Indoor Robotics[4], Embodied AI Vehicular[15]).

A particularly active line of work centers on defining and measuring perceptual quality in ways that align with embodied task performance, contrasting traditional image quality metrics with task-driven assessments. Image Quality Embodied AI[0] sits squarely within the Quality Assessment Frameworks and Benchmarks branch, specifically under Embodied-Specific Quality Assessment, where it joins efforts like Perceptual Quality Embodied[1] and RGC-VQA[49] in developing metrics that account for agent-centric visual demands. While Perceptual Quality Embodied[1] emphasizes human-aligned perceptual measures, Image Quality Embodied AI[0] appears to focus more directly on how image degradation affects downstream embodied task success, bridging quality assessment with navigation and manipulation outcomes. This contrasts with broader embodied AI surveys (e.g., Embodied AI Survey[14]) that catalog task types without deep dives into quality metrics, and with works like Embodied Image Compression[41] that optimize compression for embodied scenarios. The central tension across these branches involves balancing perceptual fidelity, computational efficiency, and task-specific relevance—questions that remain open as embodied systems scale to more complex, real-world environments.

Related Works in Same Category

The following **3 sibling papers** share the same taxonomy leaf node with the original paper:

1. Perceptual Quality Assessment for Embodied AI

Authors: C Li, J Xiao, J Zhang, F Wen, Z Zhang, et al. (6 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

â , Embodied AI collects signals â Embodied AI should be analyzed to filter out these low-quality images. For human viewers, this problem can be solved through Image Quality Assessment â

⚠ Similarity Notice

These papers share nearly identical titles, abstracts, and core technical content, including the same perception-cognition-decision-execution pipeline, the same Embodied-IQA database with over 30k image pairs and 5m annotations, and identical experimental setup with 15 IQA methods and 1.5k real-world tasks. The candidate paper appears to be a published or revised version of the original submission, with author names revealed and a project page added.

2. Embodied Image Quality Assessment for Robotic Intelligence

Authors: Zhang Jianbo, LI Chun-Yi, Hao Jie, Jia Jun, Duan Hui-yu, et al. (9 authors total) | **Year/Venue**: 2024 • arXiv.org | **URL**: View paper

Abstract

Image Quality Assessment (IQA) of User-Generated Content (UGC) is a critical technique for human Quality of Experience (QoE). However, does the the image quality of Robot-Generated Content (RGC) demonstrate traits consistent with the Moravec paradox, potentially conflicting with human perceptual norms? Human subjective scoring is more based on the attractiveness of the image. Embodied agent are required to interact and perceive in the environment, and finally perform specific tasks. Visual image...

Relationship Analysis

Both papers belong to the Embodied-Specific Quality Assessment category, focusing on developing quality metrics tailored to robot perception and embodied agent requirements rather than human preferences. They overlap in establishing databases with robot-generated content annotations and proposing IQA methods for embodied AI tasks, both addressing the gap between human visual systems and robot visual systems. The key difference is that the original paper constructs a comprehensive perception-cognition-decision-execution pipeline with over 30k image pairs and 5M annotations from VLMs/VLAs/real robots across all downstream steps, while the candidate paper focuses on a smaller 12,500-image Embodied Preference Database (EPD) and proposes a specific no-reference model (MA-EIQA) with multi-scale attention for embodied image quality assessment.

3. RGC-VQA: An Exploration Database for Robotic-Generated Video Quality Assessment

Authors: Jin Jianing, Ying Jiangyong, Duan Hui-yu, Yang Liu, Wu, et al. (11 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

As camera-equipped robotic platforms become increasingly integrated into daily life, robotic-generated videos have begun to appear on streaming media platforms, enabling us to envision a future where humans and robots coexist. We innovatively propose the concept of Robotic-Generated Content (RGC) to term these videos generated from egocentric perspective of robots. The perceptual quality of RGC videos is critical in human-robot interaction scenarios, and RGC videos exhibit unique distortions and...

Relationship Analysis

Both papers belong to the Embodied-Specific Quality Assessment category, establishing quality metrics and databases tailored to robot-generated content. They overlap in addressing perceptual quality assessment for embodied AI systems, with both creating large-scale databases with subjective annotations and benchmarking existing IQA/VQA methods. The key difference is that the original paper focuses on image quality assessment across the full perception-cognition-decision-execution pipeline using VLM/VLA annotations and real-world robot experiments, while the candidate paper specifically targets video quality assessment from robotic egocentric perspectives (drones, wheeled robots, robotic arms) using human subjective ratings and traditional VQA model benchmarking.

Contributions Analysis

Overall novelty summary. The paper proposes a perception-cognition-decision-execution pipeline for assessing image quality in embodied AI contexts, establishes the Embodied-IQA database with over 30,000 image pairs and 5 million annotations from vision-language models and real robots, and benchmarks mainstream IQA methods on this data. Within the taxonomy, it resides in the 'Embodied-Specific Quality Assessment' leaf under 'Quality Assessment Frameworks and Benchmarks', alongside three sibling papers. This leaf represents a relatively sparse research direction within a 50-paper taxonomy spanning 23 leaf nodes, suggesting the work addresses an emerging rather than saturated area.

The taxonomy reveals that quality assessment for embodied AI sits at the intersection of multiple research streams. Neighboring leaves include 'World Model and Generative Content Evaluation' (assessing scene quality and physical plausibility in generative systems) and 'General Visual Quality Assessment' (broader multimedia quality metrics). The paper's focus on robot-centric usability distinguishes it from general visual quality work, while its emphasis on task-driven metrics connects to navigation and manipulation branches. The taxonomy's scope notes clarify that embodied-specific quality assessment excludes general multimedia metrics, positioning this work as bridging perceptual quality and downstream task performance.

Among 24 candidates examined across three contributions, the analysis found limited prior work overlap. The perception-cognition-decision-execution pipeline examined 10 candidates with 1 potential refutation; the database construction examined 4 candidates with 1 refutation; and the benchmark evaluation examined 10 candidates with 2 refutations. These statistics suggest that within the top-24 semantic matches, most contributions appear relatively novel, though the search scope is modest. The pipeline and database contributions show particularly sparse prior work, while the benchmarking component encounters slightly more existing evaluation efforts.

Based on this limited search of 24 candidates, the work appears to occupy a relatively underexplored niche at the intersection of image quality assessment and embodied task performance. The sparse sibling count and low refutation rates suggest novelty, though the analysis does not cover exhaustive literature review or domain-specific venues. The taxonomy structure indicates this is an emerging research direction rather than a mature subfield, consistent with the observed scarcity of directly comparable prior work.

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

Contribution 1: Perception-cognition-decision-execution pipeline for Embodied AI quality assessment

Description: The authors develop a theoretical framework grounded in Mertonian systems and meta-cognitive theory that structures Embodied AI evaluation into four stages: perception, cognition, decision, and execution. This pipeline defines how to collect quality scores for robotic tasks.

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. Perceptual Quality Assessment for Embodied AI

URL: View paper

Prior Art Analysis

Perceptual Quality Embodied[1] demonstrates that a perception-cognition-decision-execution pipeline for Embodied AI evaluation was already established prior to the original paper. The candidate paper explicitly states they 'constructed a perceptioncognition-decision-execution pipeline' based on Mertonian systems and meta-cognitive theory, with the same four-stage structure. Both papers use this identical framework to evaluate embodied AI quality, with the candidate providing over 5 million annotations across these stages. The candidate paper also provides the theoretical foundation (Mertonian systems) that the original paper claims to have developed.

Evidence

Evidence 1 - **Rationale**: This verbatim match of theoretical explanation demonstrates that the Mertonian systems framework and its application to distinguishing robot visual systems from human/machine systems was already established in the candidate paper, undermining claims of theoretical novelty. - **Original**: hvs and mvs can both be simplified as newtonian systems, since their decision and execution processes are robust. however, the decision and execution of rvs do not fully match cognition. - **Candidate**: hvs and mvs can both be simplified as newtonian systems, since their decision and execution processes are robust. however, the decision and execution of rvs do not fully match cognition.

---

2. Embodied intelligence-based perception, decision-making, and control for autonomous operations of rail transportation

URL: View paper

Brief Assessment

Rail Transportation Embodied[57] focuses on rail transportation systems using colored Petri nets for safety evaluation. The candidate discusses embodied perception and execution in a transportation context, not a general framework for Embodied AI quality assessment across robotic tasks.

---

3. Thinking and Moving: An Efficient Computing Approach for Integrated Task and Motion Planning in Cooperative Embodied AI Systems (Invited Paper)

URL: View paper

Brief Assessment

Thinking and Moving[59] proposes a cognitive-inspired modular framework for cooperative embodied AI systems focused on task execution efficiency, not quality assessment. The original paper develops a theoretical framework for evaluating robotic task quality through perception-cognition-decision-execution stages, while the candidate addresses computational optimization in multi-agent collaboration.

---

4. Large model empowered embodied ai: A survey on decision-making and embodied learning

URL: View paper

Brief Assessment

Large Model Embodied[56] is a survey paper on decision-making and embodied learning in embodied AI, not focused on quality assessment frameworks or evaluation pipelines for robotic tasks.

---

5. Embodied AI with Large Language Models: A Survey and New HRI Framework

URL: View paper

Brief Assessment

Embodied AI LLMs[61] proposes a perception-action loop (PALoop) framework for human-robot interaction, not a quality assessment pipeline. The candidate focuses on emotional logic engines and companion robots, while the original develops an IQA evaluation framework with four distinct stages for robotic task quality scoring.

---

6. Agentic Robot: A Brain-Inspired Framework for Vision-Language-Action Models in Embodied Agents

URL: View paper

Brief Assessment

Agentic Robot[54] focuses on robotic manipulation through a perception-reasoning-execution-verification loop for task execution, not on quality assessment or evaluation frameworks for Embodied AI systems. The pipeline serves a different purpose (task control vs. quality evaluation).

---

7. Embodied Question Answering

URL: View paper

Brief Assessment

Embodied Question Answering[26] focuses on question answering through navigation in 3D environments, not on developing a quality assessment framework with perception-cognition-decision-execution stages for evaluating robotic task performance.

---

8. Robobench: A Comprehensive Evaluation Benchmark for Multimodal Large Language Models as Embodied Brain

URL: View paper

Brief Assessment

Robobench[60] focuses on evaluating MLLMs as embodied brains across five cognitive dimensions (instruction comprehension, perception reasoning, planning, affordance prediction, failure analysis) for manipulation tasks, not on developing a quality assessment pipeline for perception-cognition-decision-execution stages as proposed in the original paper.

---

9. Embodied ai agents: Modeling the world

URL: View paper

Brief Assessment

Embodied AI Modeling[58] focuses on world modeling architectures for embodied agents (virtual, wearable, robotic) rather than image quality assessment pipelines. The paper does not address quality evaluation frameworks or the specific perception-cognition-decision-execution structure proposed for IQA tasks.

---

10. A survey on vision-language-action models for embodied ai

URL: View paper

Brief Assessment

Vision Language Action Survey[55] focuses on vision-language-action models for robotic control, not on image quality assessment frameworks. The survey discusses control policies and task planners but does not address quality evaluation pipelines grounded in Mertonian systems or meta-cognitive theory.

Contribution 2: Embodied-IQA database with multi-stage annotations

Description: The authors create a large-scale database of reference and distorted image pairs for embodied tasks, annotated by VLMs, VLAs, and real robots. This resource provides fine-grained labels across cognition, decision, and execution stages to support quality metric development.

This contribution was assessed against **4 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Embodied-R1: Reinforced Embodied Reasoning for General Robotic Manipulation

URL: View paper

Brief Assessment

Embodied-R1[53] focuses on robotic manipulation through pointing-based reasoning and does not address image quality assessment databases or multi-stage quality annotations for embodied AI tasks.

---

2. Embodied Image Quality Assessment for Robotic Intelligence

URL: View paper

Prior Art Analysis

Embodied Image Quality[12] demonstrates prior work on creating an image quality assessment database specifically for embodied AI tasks. The candidate paper presents the EPD (Embodied Preference Database) containing 12,500 distorted image annotations for robot-generated content, establishing assessment metrics based on downstream robot tasks. This shows that the concept of creating specialized IQA databases with annotations for embodied AI applications existed before the original paper's claimed contribution of the Embodied-IQA database with multi-stage annotations from VLMs, VLAs, and real robots.

Evidence

Evidence 1 - **Rationale**: Both papers claim to establish databases for embodied AI image quality assessment with annotations based on robot tasks. The candidate paper's EPD predates the original's Embodied-IQA and demonstrates the concept of creating specialized databases with distorted images and robot-task-based annotations, challenging the novelty of the original's database contribution. - **Original**: we (1) based on the mertonian system and meta-cognitive theory, constructed a perceptioncognition-decision-execution pipeline and defined a comprehensive subjective score collection process; (2) established the embodied-iqa database, containing over 30k reference/distorted image pairs, with more tha... - **Candidate**: we propose the first embodied preference database (epd), which contains 12,500 distorted image annotations. we establish assessment metrics based on the downstream tasks of robot.

Evidence 2 - **Rationale**: The candidate paper explicitly addresses the distinction between human-oriented IQA and robot-oriented IQA, establishing the conceptual foundation for creating databases that assess image quality from an embodied agent's perspective rather than human perception, which is central to the original paper's claimed contribution. - **Original**: data: we add corruption to images in embodied tasks, collecting over 36k reference/distorted image pairs. we perform inference using vision language models (vlm) and vision language action-model (vla) for over 5 million annotations. - **Candidate**: image quality assessment (iqa) of user-generated content (ugc) is a critical technique for human quality of experience (qoe). however, does the the image quality of robot-generated content (rgc) demonstrate traits consistent with the moravec paradox, potentially conflicting with human perceptual nor...

---

3. Point-It-Out: Benchmarking Embodied Reasoning for Vision Language Models in Multi-Stage Visual Grounding

URL: View paper

Brief Assessment

Point-It-Out[51] focuses on visual grounding for vision-language models in embodied reasoning tasks (object localization, task-driven pointing, trajectory prediction), not on image quality assessment databases with distortion-based annotations for embodied AI.

---

4. A survey of embodied ai in healthcare: Techniques, applications, and opportunities

URL: View paper

Brief Assessment

This candidate is a survey paper on embodied AI in healthcare, focusing on medical applications. It does not present a database for image quality assessment with multi-stage annotations for embodied tasks, which is the core contribution of the original paper.

Contribution 3: Benchmark evaluation of IQA methods for Embodied AI

Description: The authors evaluate 15 existing IQA methods on their Embodied-IQA database, showing that current approaches are insufficient for robotic perception tasks. They also conduct real-world robot experiments to reveal connections among cognition, decision, and execution.

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. Improved no-reference image quality assessment algorithm based on visual perception characteristics

URL: View paper

Brief Assessment

No-Reference Image Quality[62] focuses on general no-reference image quality assessment using visual perception characteristics for standard datasets (LIVE, TID2013, KADID10K), not on benchmarking IQA methods specifically for embodied AI robotic perception tasks or evaluating connections among cognition, decision, and execution in robotic systems.

---

2. MEFormer: Enhancing Low-Light Images While Preserving Image Authenticity in Mining Environments

URL: View paper

Brief Assessment

MEFormer[67] focuses on low-light image enhancement for mining environments, not on benchmarking image quality assessment methods for embodied AI or robotic perception tasks.

---

3. Research Progress on Color Image Quality Assessment

URL: View paper

Brief Assessment

Color Image Quality[64] is a review paper on color image quality assessment methods across various domains. It does not present a benchmark evaluation specifically for embodied AI robotic perception tasks, which is the novel contribution claimed by the original paper.

---

4. Fast underwater image enhancement for improved visual perception

URL: View paper

Brief Assessment

Fast Underwater Enhancement[63] focuses on underwater image enhancement for robotic visual perception in marine environments, not on benchmarking IQA methods across diverse embodied AI tasks or establishing evaluation frameworks for robotic perception quality assessment.

---

5. Ewmbench: Evaluating scene, motion, and semantic quality in embodied world models

URL: View paper

Brief Assessment

Ewmbench[2] focuses on evaluating embodied world models for video generation in robotic manipulation tasks, not on benchmarking image quality assessment methods for robotic perception as described in the original contribution.

---

6. Embodied Image Quality Assessment for Robotic Intelligence

URL: View paper

Prior Art Analysis

Embodied Image Quality[12] demonstrates prior work on benchmarking IQA methods specifically for embodied AI applications. The candidate paper explicitly states that it verifies 'the performance of mainstream iqa algorithms on epd dataset' and proposes a novel IQA model (MA-EIQA) designed for embodied robots. This shows that evaluating existing IQA methods on embodied AI tasks and demonstrating their insufficiency was already explored before the original paper's claimed contribution.

Evidence

Evidence 1 - **Rationale**: Both papers benchmark existing IQA methods on embodied AI databases and demonstrate that current approaches are insufficient. The candidate paper's verification of mainstream IQA algorithms on the EPD dataset and demonstration that embodied image quality assessment differs from human assessment establishes prior work in this area. - **Original**: experiment: we experiment with 15 advanced iqa methods on our database, proving that more sophisticated iqa metrics are needed for embodied ai. additionally, we first conduct real-world experiments in the iqa field, executing 1.5k embodied tasks in the real-world, revealing the internal connections ... - **Candidate**: finally, the performance of mainstream iqa algorithms on epd dataset is verified. the experiments demonstrate that quality assessment of embodied images is different from that of humans.

Evidence 2 - **Rationale**: Both papers share the same goal of promoting embodied AI development through IQA evaluation, indicating that the candidate paper already established the research direction of benchmarking IQA methods for embodied AI applications. - **Original**: we sincerely hope that through evaluation, we can promote the application of embodied ai under complex distortions in the real-world. - **Candidate**: we sincerely hope that the epd can contribute to the development of embodied ai by focusing on image quality assessment.

---

7. NaviTrace: Evaluating Embodied Navigation of Vision-Language Models

URL: View paper

Brief Assessment

NaviTrace[65] focuses on evaluating vision-language models for robotic navigation tasks using visual question answering, not on image quality assessment methods for robotic perception. The technical focus is fundamentally different from IQA evaluation.

---

8. Perceptual Quality Assessment for Embodied AI

URL: View paper

Prior Art Analysis

Perceptual Quality Embodied[1] demonstrates prior work on benchmarking IQA methods for embodied AI with robotic perception tasks. The candidate paper explicitly states they 'experiment with 15 advanced iqa methods on our database, proving that more sophisticated iqa metrics are needed for embodied ai' and 'conduct real-world experiments in the iqa field, executing 1.5k embodied tasks in the real-world, revealing the internal connections between cognition, decision, and execution.' This directly corresponds to the original paper's claimed contribution of evaluating 15 IQA methods and conducting real-world robot experiments.

Evidence

Evidence 1 - **Rationale**: The identical database specifications (30k+ image pairs, 5m+ annotations from VLMs/VLAs/robots) demonstrate that the candidate paper already established the benchmark infrastructure claimed as novel by the original paper. - **Original**: established the embodied-iqa database, containing over 30k reference/distorted image pairs, with more than 5m fine-grained annotations provided by vision language models/vision language action-models/real-world robots - **Candidate**: established the embodied-iqa database, containing over 30k reference/distorted image pairs, with more than 5m fine-grained annotations provided by vision language models/vision language action-models/real-world robots

---

9. An exploration of embodied visual exploration

URL: View paper

Brief Assessment

Embodied Visual Exploration[66] focuses on visual exploration strategies for robots in 3D environments, not on benchmarking image quality assessment methods for robotic perception tasks.

---

10. Evaluating Uncertainty and Quality of Visual Language Action-enabled Robots

URL: View paper

Brief Assessment

Uncertainty VLA Robots[10] focuses on uncertainty and quality metrics for VLA models in robotic manipulation, not on benchmarking IQA methods for embodied AI perception tasks. The candidate evaluates model confidence and task execution quality, while the original evaluates image quality assessment methods.

Appendix: Text Similarity Detection

Textual similarity detection checked 23 papers and found 3 similarity segment(s) across 1 paper(s).

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

1. Perceptual Quality Assessment for Embodied AI

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

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

References

- [0] Image Quality Assessment for Embodied AI View paper
- [1] Perceptual Quality Assessment for Embodied AI View paper
- [2] Ewmbench: Evaluating scene, motion, and semantic quality in embodied world models View paper
- [3] Where are we in the search for an artificial visual cortex for embodied intelligence? View paper
- [4] Multimodal Perception for Indoor Mobile Robotics Navigation and Safe Manipulation View paper
- [5] Objectnav revisited: On evaluation of embodied agents navigating to objects View paper
- [6] Elevating Visual Perception in Multimodal LLMs with Visual Embedding Distillation View paper
- [7] Vision Language Action Models in Robotic Manipulation: A Systematic Review View paper
- [8] Card: Cross-modal agent framework for generative and editable residential design View paper
- [9] Habitat-matterport 3d dataset (hm3d): 1000 large-scale 3d environments for embodied ai View paper
- [10] Evaluating Uncertainty and Quality of Visual Language Action-enabled Robots View paper
- [11] Large Language Models for Robotics: Opportunities, Challenges, and Perspectives View paper
- [12] Embodied Image Quality Assessment for Robotic Intelligence View paper
- [13] Affordance-Based Goal Imagination for Embodied AI Agents View paper
- [14] A survey of embodied ai: From simulators to research tasks View paper
- [15] Embodied AI-Enhanced Vehicular Networks: An Integrated Large Language Models and Reinforcement Learning Method View paper
- [16] Multi-Step Guided Diffusion for Image Restoration on Edge Devices: Toward Lightweight Perception in Embodied AI View paper
- [17] Generative physical ai in vision: A survey View paper
- [18] Are Multimodal Large Language Models Ready for Omnidirectional Spatial Reasoning? View paper
- [19] World-in-World: World Models in a Closed-Loop World View paper
- [20] GSArch: Breaking Memory Barriers in 3D Gaussian Splatting Training via Architectural Support View paper
- [21] Generic Simulation Framework for Evaluation Process: Applied to AI-powered Visual Perception System in Autonomous Driving View paper
- [22] No Redundancy, No Stall: Lightweight Streaming 3D Gaussian Splatting for Real-time Rendering View paper
- [23] Are Video Models Ready as Zero-Shot Reasoners? An Empirical Study with the MME-CoF Benchmark View paper
- [24] Generative world explorer View paper
- [25] TextToucher: Fine-Grained Text-to-Touch Generation View paper
- [26] Embodied Question Answering View paper
- [27] VidEgoThink: Assessing Egocentric Video Understanding Capabilities for Embodied AI View paper
- [28] Perceptual Visual Quality Assessment in Multimedia Communication View paper
- [29] A control-centric benchmark for video prediction View paper
- [30] A lightweight and style-robust neural network for autonomous driving in end side devices View paper
- [31] B-AIS: An Automated Process for Black-box Evaluation of Visual Perception in AI-enabled Software against Domain Semantics View paper
- [32] Shadowplay: An Embodied AI Art Installation View paper
- [33] Multi-view Enhanced Few-shot Action Recognition for Embodied Visual Perception View paper
- [34] A Survey of Object Goal Navigation: Datasets, Metrics and Methods View paper
- [35] RAI: Flexible Agent Framework for Embodied AI View paper
- [36] Optical performance perception technology in the embodied intelligence-based optical astronomical telescopes View paper
- [37] AirCopBench: A Benchmark for Multi-drone Collaborative Embodied Perception and Reasoning View paper
- [38] TwinOR: Photorealistic Digital Twins of Dynamic Operating Rooms for Embodied AI Research View paper
- [39] PerTouch: VLM-Driven Agent for Personalized and Semantic Image Retouching View paper
- [40] RynnEC: Bringing MLLMs into Embodied World View paper
- [41] Embodied Image Compression View paper
- [42] RAISECity: A Multimodal Agent Framework for Reality-Aligned 3D World Generation at City-Scale View paper
- [43] Image Quality Assessment in Visual Reinforcement Learning for Fast-moving Targets View paper
- [44] Perception Matters: Enhancing Embodied AI with Uncertainty-Aware Semantic Segmentation View paper

- [45] XGC-AVis: Towards Audio-Visual Content Understanding with a Multi-Agent Collaborative System View paper
- [46] OGGSplat: Open Gaussian Growing for Generalizable Reconstruction with Expanded Field-of-View View paper
- [47] Skip-SCAR: Hardware-Friendly High-Quality Embodied Visual Navigation View paper
- [48] Wanderland: Geometrically Grounded Simulation for Open-World Embodied AI View paper
- [49] RGC-VQA: An Exploration Database for Robotic-Generated Video Quality Assessment View paper
- [50] CC-FMO: Camera-Conditioned Zero-Shot Single Image to 3D Scene Generation with Foundation Model Orchestration View paper
- [51] Point-It-Out: Benchmarking Embodied Reasoning for Vision Language Models in Multi-Stage Visual Grounding View paper
- [52] A survey of embodied ai in healthcare: Techniques, applications, and opportunities View paper
- [53] Embodied-R1: Reinforced Embodied Reasoning for General Robotic Manipulation View paper
- [54] Agentic Robot: A Brain-Inspired Framework for Vision-Language-Action Models in Embodied Agents View paper
- [55] A survey on vision-language-action models for embodied ai View paper
- [56] Large model empowered embodied ai: A survey on decision-making and embodied learning View paper
- [57] Embodied intelligence-based perception, decision-making, and control for autonomous operations of rail transportation View paper
- [58] Embodied ai agents: Modeling the world View paper
- [59] Thinking and Moving: An Efficient Computing Approach for Integrated Task and Motion Planning in Cooperative Embodied AI Systems (Invited Paper) View paper
- [60] Robobench: A Comprehensive Evaluation Benchmark for Multimodal Large Language Models as Embodied Brain View paper
- [61] Embodied AI with Large Language Models: A Survey and New HRI Framework View paper
- [62] Improved no-reference image quality assessment algorithm based on visual perception characteristics View paper
- [63] Fast underwater image enhancement for improved visual perception View paper
- [64] Research Progress on Color Image Quality Assessment View paper
- [65] NaviTrace: Evaluating Embodied Navigation of Vision-Language Models View paper
- [66] An exploration of embodied visual exploration View paper
- [67] MEFormer: Enhancing Low-Light Images While Preserving Image Authenticity in Mining Environments View paper