Novelty Assessment Report

Paper: Spatial CAPTCHA: Generatively Benchmarking Spatial Reasoning for Human-Machine Differentiation

PDF URL: https://openreview.net/pdf?id=8LAc5kzjhz

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2026-01-08

Abstract

Online services rely on CAPTCHAs as a first line of defense against automated abuse, yet recent advances in multi-modal large language models (MLLMs) have eroded the effectiveness of conventional designs that focus on text recognition or 2D image understanding. To address this challenge, we present **Spatial CAPTCHA**, a novel human-verification framework that leverages fundamental differences in spatial reasoning between humans and MLLMs. Unlike existing CAPTCHAs that rely on low-level perception tasks vulnerable to modern AI, Spatial CAPTCHA generates dynamic questions requiring geometric reasoning, perspective-taking, occlusion handling, and mental rotation—skills intuitive for humans but difficult for current AI systems. The system employs a procedural generation pipeline with constraint-based difficulty control, automated correctness verification, and human-in-the-loop validation to ensure scalability, robustness, and adaptability. Evaluation on a corresponding benchmark, **Spatial-CAPTCHA-Bench**, demonstrates that humans vastly outperform 10 state-of-the-art MLLMs, with the best model achieving only 31.0\% Pass@1 accuracy. Result comparison with Google reCAPTCHA further confirms the effectiveness of Spatial CAPTCHA as both a security mechanism and a diagnostic tool for spatial reasoning in AI.

Core Task Landscape

This paper addresses: **Spatial Reasoning for Human-Machine Differentiation**

A total of **36 papers** were analyzed and organized into a taxonomy with **18 categories**.

Taxonomy Overview

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

- **Spatial Reasoning Benchmarks and Evaluation Frameworks**
- **Methods for Enhancing Spatial Reasoning in AI Systems**
- **Human-Machine Differentiation and Security Applications**
- **Application Domains and Real-World Integration**
- **Cognitive Foundations and Theoretical Perspectives**
- **Modality-Specific Spatial Reasoning Challenges**

Complete Taxonomy Tree

- Spatial Reasoning for Human-Machine Differentiation Survey Taxonomy
- Spatial Reasoning Benchmarks and Evaluation Frameworks
  - General Spatial Relation Understanding Benchmarks (4 papers)
  - [1] Visual Spatial Reasoning (Liu, 2023) View paper
  - [2] What's "up" with vision-language models? Investigating their struggle with spatial reasoning (Chang, 2023) View paper
  - [3] Is A Picture Worth A Thousand Words? Delving Into Spatial Reasoning for Vision Language Models (Neel Joshi, 2024) View paper
  - [14] Expand VSR Benchmark for VLLM to Expertize in Spatial Rules (Liu Bing-quan, 2024) View paper
  - Hierarchical and Multi-Dimensional Spatial Evaluation (3 papers)
  - [5] 11plus-bench: Demystifying multimodal llm spatial reasoning with cognitive-inspired analysis (Li, 2025) View paper
  - [6] Sphere: Unveiling spatial blind spots in vision-language models through hierarchical evaluation (Koenecke, 2025) View paper
  - [28] Spatial Reasoning in Multimodal Large Language Models: A Survey of Tasks, Benchmarks and Methods (Weichen Liu, 2025) View paper
  - Specialized Spatial Reasoning Tasks (4 papers)
  - [4] TopViewRS: Vision-Language Models as Top-View Spatial Reasoners (Collier, 2024) View paper
  - [7] Sqa3d: Situated question answering in 3d scenes (Xinbin Ma, 2022) View paper
  - [24] GamiBench: Evaluating Spatial Reasoning and 2D-to-3D Planning Capabilities of MLLMs with Origami Folding Tasks (Ryan Spencer, 2025) View paper
  - [27] RotBench: Evaluating Multimodal Large Language Models on Identifying Image Rotation (Cho, 2025) View paper
  - Qualitative and Simulation-Based Spatial Reasoning (2 papers)
  - [9] Reframing spatial reasoning evaluation in language models: A real-world simulation benchmark for qualitative reasoning (Li, 2024) View paper
  - [30] How Far are VLMs from Visual Spatial Intelligence? A Benchmark-Driven Perspective (Yu Songsong, 2025) View paper
- Methods for Enhancing Spatial Reasoning in AI Systems
  - Chain-of-Thought and Reasoning Process Enhancement (2 papers)
  - [16] Improving Spatial Reasoning in Vision-Language Models via Chain-of-Thought Annotation and Reinforcement Learning (T Huang, 2025) View paper
  - [22] Reasoning under Vision: Understanding Visual-Spatial Cognition in Vision-Language Models for CAPTCHA (Python Song, 2025) View paper

- Imagination and Internal Simulation Mechanisms (2 papers)
- [11] Jigsaw-Puzzles: From Seeing to Understanding to Reasoning in Vision-Language Models (Zhang, 2025) View paper
- [21] Imagine in Space: Exploring the Frontier of Spatial Intelligence and Reasoning Efficiency in Vision Language Models (Xiaoxing Lian, 2025) View paper
- Physics-Informed and Human-Prior Integration (2 papers)
- [10] Mind meets space: Rethinking agentic spatial intelligence from a neuroscience-inspired perspective (ZHANG Zetong, 2025) View paper
- [23] Towards Physics-informed Spatial Intelligence with Human Priors: An Autonomous Driving Pilot Study (Wu, 2025) View paper

- Human-Machine Differentiation and Security Applications
  - Spatial CAPTCHA and Verification Systems ★ (1 papers)
  - [0] Spatial CAPTCHA: Generatively Benchmarking Spatial Reasoning for Human-Machine Differentiation (Anon et al., 2026) View paper
  - Human-Machine Performance Comparison Studies (2 papers)
  - [18] Spatial Heuristics and Random Spatial Exploration: Children, Adults and the Machine Coloring-in Places in The Grid Game (Christiane Langeâ□□KÃ¼ttner, 2025) View paper
  - [19] Comparing machines and humans on a visual categorization test (FranÃ§ois Fleuret, 2011) View paper
  - Object Authenticity and Visual Discrimination (1 papers)
  - [12] Exploring the Limits of Large Language Modelsâ□□ Ability to Distinguish Between Objects (Hyeongjin Ju, 2025) View paper
- Application Domains and Real-World Integration
  - Autonomous Driving and Navigation (1 papers)
  - [8] Vision-Integrated LLMs for Autonomous Driving Assistance: Human Performance Comparison and Trust Evaluation (Kim Nam-Hee, 2025) View paper
  - Human-Robot Spatial Collaboration (3 papers)
  - [13] Spatial reasoning for human robot interaction (E. A. Sisbot, 2007) View paper
  - [33] Recent development of human-robot natural interaction in spatial cognition tasks (Xianliang Mu, 2016) View paper
  - [34] Spatial representation and reasoning for human-robot collaboration (William G. Kennedy, 2007) View paper
  - Interior Design and Space Planning (1 papers)
  - [20] Enhancing interior design and space planning via humanâ□□machine intelligent interaction for artistic cognition (Jiatong Jiang, 2025) View paper
- Cognitive Foundations and Theoretical Perspectives
  - Human Cognitive Mechanisms and Development (1 papers)
  - [15] Language and the development of spatial reasoning (Anna Shusterman, 2005) View paper
  - Neuroscience-Inspired and Embodied Cognition (2 papers)
  - [17] Neurophysiological and Behavioral Differences in Human-Multiagent Tasks: An EEG Network Perspective (Gregory Bales, 2022) View paper
  - [31] Embodied spatial cognition: Biological and artificial systems (H. Mallot, 2009) View paper
  - Comparative Cognition and AI-Human Differences (1 papers)
  - [25] Why Multimodal Models Struggle with Spatial Reasoning: Insights from Human Cognition (Scott, 2025) View paper
  - Qualitative Spatial Reasoning Theory (2 papers)
  - [26] Some Connections between Qualitative Spatial Reasoning and Machine Learning (Cohn, 2024) View paper
  - [32] Qualitative spatial reasoning: The CLOCK project (Kenneth D. Forbus, 1991) View paper
- Modality-Specific Spatial Reasoning Challenges (3 papers)
  - [29] Spatial Blind Spot: Auditory Motion Perception Deficits in Audio LLMs (Zhe Sun, 2025) View paper
  - [35] Minimal videos: Trade-off between spatial and temporal information in human and machine vision. (Guy Ben-Yosef, 2021) View paper
  - [36] Exploring the spatial reasoning ability of neural models in human IQ tests. (Kim HyunJae, 2021) View paper

Narrative

Core task: spatial reasoning for human-machine differentiation. The field encompasses diverse branches that collectively address how AI systems understand and manipulate spatial information, and how these capabilities can distinguish human from machine cognition. The taxonomy reveals several major directions: benchmarks and evaluation frameworks that systematically test spatial abilities across modalities; methods for enhancing spatial reasoning through architectural innovations, training strategies, and cognitive-inspired approaches; human-machine differentiation and security applications that exploit spatial reasoning gaps; application domains spanning robotics, autonomous driving, and interactive systems; cognitive foundations drawing from developmental psychology and neuroscience; and modality-specific challenges in vision, language, and embodied settings. Representative works like Visual Spatial Reasoning[1] and Whats Up VLMs[2] illustrate benchmark development, while Picture Worth Spatial[3] and TopViewRS[4] explore domain-specific spatial understanding.

Recent activity highlights contrasting themes: some lines pursue comprehensive benchmarking of vision-language models on spatial tasks (e.g., 11plus bench[5], Sphere Blind Spots[6]), revealing persistent weaknesses in rotation, orientation, and relational reasoning, while others investigate cognitive and theoretical underpinnings (Reframing Spatial Reasoning[9], Mind Meets Space[10]) to inform better architectures. The original paper, Spatial CAPTCHA[0], sits squarely within the human-machine differentiation branch alongside Vision CAPTCHA Reasoning[22], leveraging spatial reasoning challenges as verification mechanisms. Unlike broader benchmarks that assess general spatial competence, this work focuses on exploiting the gap between human intuitive spatial processing and current AI limitations for security purposes. This positions it as a practical application of observed weaknesses, contrasting with efforts like CoT Spatial Reasoning[16] or Multimodal Spatial Struggle[25] that aim to close the performance gap through improved reasoning strategies.

Related Works in Same Category

No sibling papers were found in the same taxonomy leaf. A taxonomy-subtopic-level comparison will be produced instead.

Taxonomy-Level Summary

The original leaf focuses specifically on CAPTCHA and verification systems that use spatial reasoning challenges to differentiate humans from machines. The sibling subtopics address related but distinct aspects: one examines empirical performance comparisons between humans and machines on cognitive tasks, while the other focuses on authenticity detection and visual discrimination. All three share the common thread of human-machine differentiation but approach it from different methodological angles.

Similarities: - All three subtopics involve distinguishing human capabilities from machine capabilities - All leverage spatial or visual reasoning as a core component - All address the broader challenge of understanding differences between human and machine cognition - All exclude overlapping concerns to maintain clear boundaries (e.g., CAPTCHA systems are excluded from the sibling topics)

Differences: - The original leaf is application-focused (verification systems), while siblings are research-focused (performance studies and discrimination tasks) - The original leaf emphasizes interactive challenges (geometric reasoning, perspective-taking), while siblings focus on evaluation and detection - Human-Machine Performance Comparison Studies explicitly compares capabilities empirically, while the original leaf implements challenges based on assumed capability differences - Object Authenticity focuses on detecting fake/real distinctions in visual content, while the original leaf uses spatial tasks as verification mechanisms - The original leaf has a security/verification purpose, while siblings have analytical/evaluative purposes

Suggested Search Directions: - Investigate how findings from performance comparison studies inform CAPTCHA design - Explore whether authenticity detection methods could be weaponized to break spatial CAPTCHAs - Examine the intersection of spatial reasoning challenges with adversarial robustness in verification systems

Sibling Subtopics

- **Human-Machine Performance Comparison Studies** (leaves: 1, papers: 2)
- Scope: Empirical studies comparing human and machine performance on spatial reasoning, visual categorization, or cognitive tasks.
- Exclude: Excludes CAPTCHA systems and neuroscience-based analyses; those belong under verification systems or cognitive foundations respectively.
- **Object Authenticity and Visual Discrimination** (leaves: 1, papers: 1)
- Scope: Studies evaluating models' ability to distinguish real objects from replicas, humans from human-like entities, or detect visual authenticity.
- Exclude: Excludes spatial relation understanding and CAPTCHA systems; those belong under evaluation frameworks or verification systems respectively.

Contributions Analysis

Overall novelty summary. The paper introduces Spatial CAPTCHA, a human-verification framework exploiting spatial reasoning gaps between humans and MLLMs. It resides in the 'Spatial CAPTCHA and Verification Systems' leaf under 'Human-Machine Differentiation and Security Applications', where it is currently the sole paper. This isolation suggests the work occupies a sparse, emerging research direction within the broader spatial reasoning landscape, which comprises 36 papers across diverse benchmarks, methods, and application domains. The taxonomy reveals that while spatial reasoning evaluation is well-populated, security-oriented applications leveraging these gaps remain underexplored.

The taxonomy tree shows neighboring leaves include 'Human-Machine Performance Comparison Studies' (2 papers) and 'Object Authenticity and Visual Discrimination' (1 paper), both focused on empirical performance analysis rather than security applications. Broader sibling branches address 'Spatial Reasoning Benchmarks' (13 papers across 4 leaves) and 'Methods for Enhancing Spatial Reasoning' (5 papers across 3 leaves). The original paper diverges from these directions by applying observed spatial reasoning deficits to practical verification tasks, rather than benchmarking capabilities or improving model performance. Its scope_note explicitly excludes general performance comparisons and object classification, positioning it as a security-focused application distinct from adjacent evaluation-centric work.

Among 23 candidates examined, none clearly refute the three core contributions. The 'Spatial CAPTCHA framework' contribution examined 10 candidates with zero refutable overlaps; the 'Procedural generation pipeline' examined 3 candidates with similar results; and the 'Spatial-CAPTCHA-Bench' examined 10 candidates, also finding no prior work providing overlapping benchmarks. This limited search scope suggests that within the top-K semantic matches and citation expansions analyzed, no existing work combines spatial reasoning challenges with CAPTCHA-style verification and automated generation pipelines. However, the analysis does not claim exhaustive coverage of all possible prior art in security or spatial reasoning domains.

Given the limited search scope of 23 candidates, the work appears novel in its specific application of spatial reasoning to human verification. The taxonomy context reinforces this impression: the leaf contains only this paper, and adjacent leaves focus on performance analysis rather than security. However, the analysis cannot rule out relevant prior work in broader CAPTCHA literature, adversarial robustness, or cognitive security domains not captured by the semantic search. The novelty assessment is thus conditional on the examined candidate set and the taxonomy's coverage of spatial reasoning research.

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

Contribution 1: Spatial CAPTCHA framework for human-machine differentiation

Description: The authors introduce a new CAPTCHA system that exploits the gap between human and machine spatial reasoning capabilities. The framework generates dynamic questions requiring geometric reasoning, perspective-taking, occlusion handling, and mental rotation—skills intuitive for humans but challenging for current AI systems.

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. Vrc-graphnet: A graph neural network-based reasoning framework for attacking visual reasoning captchas

URL: View paper

Brief Assessment

VRC GraphNet[47] focuses on attacking visual reasoning CAPTCHAs using graph neural networks, rather than proposing a new CAPTCHA system. The candidate paper is about breaking existing CAPTCHAs, not creating frameworks that exploit spatial reasoning differences between humans and machines.

---

2. Designing Cognitive 3D Immersive CAPTCHA for Enhancing Security of Virtual Reality Systems

URL: View paper

Brief Assessment

Cognitive 3D CAPTCHA[52] focuses on VR-specific authentication using immersive 3D interactions and cognitive tasks in virtual reality environments, not general spatial reasoning benchmarks for evaluating MLLMs across diverse task types.

---

3. MF-GGNN: Crack Visual Reasoning CAPTCHA Holistically Using a Novel Multi-Feature Fusion-Based Graph Gated Neural Network

URL: View paper

Brief Assessment

MF GGNN[51] focuses on solving visual reasoning CAPTCHAs using graph neural networks for crack detection, not on designing CAPTCHA systems that exploit spatial reasoning differences between humans and machines.

---

4. NGCaptcha: A CAPTCHA Bridging the Past and the Future

URL: View paper

Brief Assessment

NGCaptcha[49] focuses on combining proof-of-work mechanisms with illusion-based image selection tasks, rather than exploiting spatial reasoning capabilities like geometric reasoning, perspective-taking, occlusion handling, and mental rotation that are central to the original paper's framework.

---

5. A captcha design based on visual reasoning

URL: View paper

Brief Assessment

Visual Reasoning CAPTCHA[53] focuses on text-query-based object finding in images using visual reasoning, not spatial reasoning tasks like geometric reasoning, perspective-taking, or mental rotation that are central to the original paper's framework.

---

6. Image CAPTCHA: based on human understanding of real world distances

URL: View paper

Brief Assessment

Real World Distances[54] focuses on depth perception and relative distance judgments using real-world object knowledge, while the original paper presents a comprehensive framework targeting multiple spatial reasoning abilities (geometric reasoning, perspective-taking, occlusion handling, mental rotation) through procedurally generated tasks with formal invariant specifications.

---

7. Robust CAPTCHAs towards malicious OCR

URL: View paper

Brief Assessment

Robust CAPTCHAs OCR[48] focuses on character-based CAPTCHAs using adversarial perturbations to exploit OCR vulnerabilities, not spatial reasoning tasks requiring geometric understanding, perspective-taking, or 3D mental rotation as in the original paper.

---

8. Adversarial Text-Based CAPTCHA Generation Method Utilizing Spatial Smoothing

URL: View paper

Brief Assessment

Adversarial Text CAPTCHA[50] focuses on adversarial noise and spatial smoothing techniques for text-based CAPTCHAs, not spatial reasoning tasks requiring geometric understanding, perspective-taking, or 3D mental rotation that distinguish the original paper's approach.

---

9. Attacks and design of image recognition CAPTCHAs

URL: View paper

Brief Assessment

Image Recognition Attacks[55] focuses on image recognition CAPTCHAs (IRCs) that rely on object recognition and contextual understanding in 2D images. The original paper's spatial CAPTCHA targets 3D spatial reasoning, geometric transformations, perspective-taking, and mental rotation—fundamentally different cognitive abilities not addressed in the candidate.

---

10. Reasoning under Vision: Understanding Visual-Spatial Cognition in Vision-Language Models for CAPTCHA

URL: View paper

Brief Assessment

Vision CAPTCHA Reasoning[22] focuses on evaluating reasoning capabilities in existing VLMs for solving CAPTCHAs, rather than proposing a new CAPTCHA generation framework. The candidate analyzes how step-by-step reasoning improves CAPTCHA solving, while the original introduces a novel procedural generation system exploiting spatial reasoning gaps.

Contribution 2: Procedural generation pipeline with constraint-based difficulty control

Description: The authors develop an autonomous pipeline that can generate unlimited CAPTCHA instances with controlled difficulty levels. The system includes mechanisms for automated correctness verification and human validation to ensure the generated challenges are both scalable and robust for real-world deployment.

This contribution was assessed against **3 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. HiEI: A universal framework for generating high-quality emerging images from natural images

URL: View paper

Brief Assessment

HiEI Framework[44] focuses on generating emerging images (two-tone images) from natural images with perceived difficulty adjustment for CAPTCHA tasks, not on procedural generation pipelines with constraint-based difficulty control for spatial reasoning tasks as described in the original paper.

---

2. Aura-CAPTCHA: A Reinforcement Learning and GAN-Enhanced Multi-Modal CAPTCHA System

URL: View paper

Brief Assessment

Aura CAPTCHA[45] focuses on reinforcement learning-based adaptive difficulty tuning for CAPTCHA systems, not on procedural generation pipelines with constraint-based difficulty control for spatial reasoning tasks. The candidate's RL approach adjusts difficulty based on user behavior metrics (incorrect attempts, response time), while the original paper's contribution involves a comprehensive procedural generation framework with geometric constraints, automated correctness verification, and human-in-the-loop validation for spatial reasoning challenges.

---

3. Thesis Supervisor: Takeo Igarashi äº□å□□åµ□ å□¥å¤«

URL: View paper

Brief Assessment

Thesis Supervisor Igarashi[46] focuses on computational design methods for parameter tweaking in visual aesthetic preference domains (photo color enhancement, lighting design). It does not address CAPTCHA generation, procedural generation pipelines for security challenges, or constraint-based difficulty control for human-machine differentiation tasks.

Contribution 3: Spatial-CAPTCHA-Bench benchmark dataset

Description: The authors create a benchmark comprising 1050 instances across seven task formulations and four spatial-ability categories, stratified into three difficulty levels. This benchmark enables standardized offline evaluation of both human and machine spatial reasoning capabilities.

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. Reframing spatial reasoning evaluation in language models: A real-world simulation benchmark for qualitative reasoning

URL: View paper

Brief Assessment

Reframing Spatial Reasoning[9] focuses on qualitative spatial reasoning evaluation in language models using 3D simulation data with topological, directional, and distance relations. The original paper's Spatial-CAPTCHA-Bench targets human-machine differentiation through procedurally generated spatial reasoning tasks for CAPTCHA applications, representing fundamentally different evaluation purposes and methodologies.

2. MMMR: Benchmarking Massive Multi-Modal Reasoning Tasks

URL: View paper

Brief Assessment

MMMR Benchmark[37] focuses on evaluating multi-modal reasoning capabilities across six diverse reasoning types (logic, math, space-time, code, map, science) in MLLMs, not on human-machine differentiation through spatial reasoning CAPTCHAs. The original paper creates a benchmark specifically for CAPTCHA-based spatial reasoning to distinguish humans from machines in security contexts, while MMMR Benchmark[37] targets diagnostic evaluation of model reasoning quality across general cognitive tasks.

3. Spatial-DISE: A Unified Benchmark for Evaluating Spatial Reasoning in Vision-Language Models

URL: View paper

Brief Assessment

Spatial DISE[43] focuses on evaluating spatial reasoning in VLMs through cognitive science-based tasks (intrinsic/extrinsic, static/dynamic), not on human-machine differentiation for CAPTCHA security. The original paper's benchmark is designed specifically for CAPTCHA deployment with procedural generation for bot defense, while Spatial DISE[43] targets diagnostic evaluation of spatial cognition in AI models.

4. Spatial reasoning with vision-language models in ego-centric multi-view scenes

URL: View paper

Brief Assessment

Egocentric Spatial Reasoning[41] focuses on ego-centric multi-view outdoor spatial reasoning for embodied AI agents (robots, self-driving cars), not human-machine differentiation via CAPTCHA. The benchmark design, task categories, and evaluation goals are fundamentally different.

5. 11plus-bench: Demystifying multimodal llm spatial reasoning with cognitive-inspired analysis

URL: View paper

Brief Assessment

11plus-bench[5] focuses on evaluating spatial reasoning in MLLMs using cognitive science-inspired standardized tests for teenagers, while the original paper creates a CAPTCHA benchmark for human-machine differentiation in online security contexts. The datasets serve fundamentally different purposes and target different evaluation scenarios.

6. Grasp: A grid-based benchmark for evaluating commonsense spatial reasoning

URL: View paper

Brief Assessment

Grasp Benchmark[42] focuses on grid-based commonsense spatial reasoning for energy collection tasks with LLM agents, while the original paper's Spatial-CAPTCHA-Bench targets human-machine differentiation through spatial reasoning challenges across seven task formulations. The benchmarks serve fundamentally different purposes and evaluate different capabilities.

7. Reasoning under Vision: Understanding Visual-Spatial Cognition in Vision-Language Models for CAPTCHA

URL: View paper

Brief Assessment

Vision CAPTCHA Reasoning[22] introduces CAPTCHA-X with reasoning annotations for evaluating VLMs, which differs fundamentally from Spatial-CAPTCHA-Bench. The candidate focuses on reasoning step annotations and action sequences, while the original provides procedurally generated instances with difficulty stratification and spatial-ability categories.

8. Raven: A dataset for relational and analogical visual reasoning

URL: View paper

Brief Assessment

Raven Dataset[40] focuses on relational and analogical visual reasoning using Raven's Progressive Matrices (RPM), not on spatial reasoning for human-machine differentiation in CAPTCHA contexts. The datasets serve fundamentally different purposes and evaluation domains.

9. ViewSpatial-Bench: Evaluating Multi-perspective Spatial Localization in Vision-Language Models

URL: View paper

Brief Assessment

ViewSpatial-Bench[38] focuses on multi-perspective spatial localization in vision-language models across camera and human viewpoints, not on human-machine differentiation through CAPTCHA tasks with difficulty stratification.

10. Physion: Evaluating Physical Prediction from Vision in Humans and Machines

URL: View paper

Brief Assessment

Physion[39] focuses on evaluating physical prediction from vision (e.g., object collisions, dynamics, rolling motion) rather than spatial reasoning for human-machine differentiation via CAPTCHA tasks. The datasets serve fundamentally different purposes and evaluation contexts.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] Spatial CAPTCHA: Generatively Benchmarking Spatial Reasoning for Human-Machine Differentiation View paper
- [1] Visual Spatial Reasoning View paper
- [2] What's "up" with vision-language models? Investigating their struggle with spatial reasoning View paper
- [3] Is A Picture Worth A Thousand Words? Delving Into Spatial Reasoning for Vision Language Models View paper
- [4] TopViewRS: Vision-Language Models as Top-View Spatial Reasoners View paper
- [5] 11plus-bench: Demystifying multimodal llm spatial reasoning with cognitive-inspired analysis View paper
- [6] Sphere: Unveiling spatial blind spots in vision-language models through hierarchical evaluation View paper
- [7] Sqa3d: Situated question answering in 3d scenes View paper
- [8] Vision-Integrated LLMs for Autonomous Driving Assistance: Human Performance Comparison and Trust Evaluation View paper
- [9] Reframing spatial reasoning evaluation in language models: A real-world simulation benchmark for qualitative reasoning View paper
- [10] Mind meets space: Rethinking agentic spatial intelligence from a neuroscience-inspired perspective View paper
- [11] Jigsaw-Puzzles: From Seeing to Understanding to Reasoning in Vision-Language Models View paper
- [12] Exploring the Limits of Large Language Modelsâ Ability to Distinguish Between Objects View paper
- [13] Spatial reasoning for human robot interaction View paper
- [14] Expand VSR Benchmark for VLLM to Expertize in Spatial Rules View paper
- [15] Language and the development of spatial reasoning View paper
- [16] Improving Spatial Reasoning in Vision-Language Models via Chain-of-Thought Annotation and Reinforcement Learning View paper
- [17] Neurophysiological and Behavioral Differences in Human-Multiagent Tasks: An EEG Network Perspective View paper
- [18] Spatial Heuristics and Random Spatial Exploration: Children, Adults and the Machine Coloring-in Places in The Grid Game View paper
- [19] Comparing machines and humans on a visual categorization test View paper
- [20] Enhancing interior design and space planning via humanâmachine intelligent interaction for artistic cognition View paper
- [21] Imagine in Space: Exploring the Frontier of Spatial Intelligence and Reasoning Efficiency in Vision Language Models View paper
- [22] Reasoning under Vision: Understanding Visual-Spatial Cognition in Vision-Language Models for CAPTCHA View paper
- [23] Towards Physics-informed Spatial Intelligence with Human Priors: An Autonomous Driving Pilot Study View paper
- [24] GamiBench: Evaluating Spatial Reasoning and 2D-to-3D Planning Capabilities of MLLMs with Origami Folding Tasks View paper
- [25] Why Multimodal Models Struggle with Spatial Reasoning: Insights from Human Cognition View paper
- [26] Some Connections between Qualitative Spatial Reasoning and Machine Learning View paper
- [27] RotBench: Evaluating Multimodal Large Language Models on Identifying Image Rotation View paper
- [28] Spatial Reasoning in Multimodal Large Language Models: A Survey of Tasks, Benchmarks and Methods View paper
- [29] Spatial Blind Spot: Auditory Motion Perception Deficits in Audio LLMs View paper
- [30] How Far are VLMs from Visual Spatial Intelligence? A Benchmark-Driven Perspective View paper
- [31] Embodied spatial cognition: Biological and artificial systems View paper
- [32] Qualitative spatial reasoning: The CLOCK project View paper
- [33] Recent development of human-robot natural interaction in spatial cognition tasks View paper
- [34] Spatial representation and reasoning for human-robot collaboration View paper
- [35] Minimal videos: Trade-off between spatial and temporal information in human and machine vision. View paper
- [36] Exploring the spatial reasoning ability of neural models in human IQ tests. View paper
- [37] MMMR: Benchmarking Massive Multi-Modal Reasoning Tasks View paper
- [38] ViewSpatial-Bench: Evaluating Multi-perspective Spatial Localization in Vision-Language Models View paper
- [39] Physion: Evaluating Physical Prediction from Vision in Humans and Machines View paper
- [40] Raven: A dataset for relational and analogical visual reasoning View paper
- [41] Spatial reasoning with vision-language models in ego-centric multi-view scenes View paper
- [42] Grasp: A grid-based benchmark for evaluating commonsense spatial reasoning View paper
- [43] Spatial-DISE: A Unified Benchmark for Evaluating Spatial Reasoning in Vision-Language Models View paper
- [44] HiEI: A universal framework for generating high-quality emerging images from natural images View paper
- [45] Aura-CAPTCHA: A Reinforcement Learning and GAN-Enhanced Multi-Modal CAPTCHA System View paper
- [46] Thesis Supervisor: Takeo Igarashi äºåäµ å¥å¤« View paper
- [47] Vrc-graphnet: A graph neural network-based reasoning framework for attacking visual reasoning captchas View paper
- [48] Robust CAPTCHAs towards malicious OCR View paper
- [49] NGCaptcha: A CAPTCHA Bridging the Past and the Future View paper
- [50] Adversarial Text-Based CAPTCHA Generation Method Utilizing Spatial Smoothing View paper
- [51] MF-GGNN: Crack Visual Reasoning CAPTCHA Holistically Using a Novel Multi-Feature Fusion-Based Graph Gated Neural Network View paper
- [52] Designing Cognitive 3D Immersive CAPTCHA for Enhancing Security of Virtual Reality Systems View paper
- [53] A captcha design based on visual reasoning View paper
- [54] Image CAPTCHA: based on human understanding of real world distances View paper
- [55] Attacks and design of image recognition CAPTCHAs View paper