Novelty Assessment Report

Paper: Manipulation as in Simulation: Enabling Accurate Geometry Perception in Robots
PDF URL: https://openreview.net/pdf?id=sWyX1BpeN4
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2026-01-01

Abstract

Modern robotic manipulation primarily relies on visual observations in a 2D color space for skill learning but suffers from poor generalization. In contrast, humans, living in a 3D world, depend more on physical properties-such as distance, size, and shape-than on texture when interacting with objects. Since such 3D geometric information can be acquired from widely available depth cameras, it appears feasible to endow robots with similar perceptual capabilities. Our pilot study found that using depth cameras for manipulation is challenging, primarily due to their limited accuracy and susceptibility to various types of noise. In this work, we propose Camera Depth Models (CDMs) as a simple plugin on daily-use depth cameras, which take RGB images and raw depth signals as input and output denoised, accurate metric depth. To achieve this, we develop a neural data engine that generates high-quality paired data from simulation by modeling a depth camera's noise pattern. Our results show that CDMs achieve nearly simulation-level accuracy in depth prediction, effectively bridging the sim-to-real gap for manipulation tasks. Notably, our experiments demonstrate, for the first time, that a policy trained on raw simulated depth, without the need for adding noise or real-world fine-tuning, generalizes seamlessly to real-world robots on two challenging long-horizon tasks involving articulated, reflective, and slender objects, with little to no performance degradation. We hope our findings will inspire future research in utilizing simulation data and 3D information in general robot policies.

Core Task Landscape

This paper addresses: **Depth Enhancement for Robotic Manipulation Using Camera Depth Models**

A total of **50 papers** were analyzed and organized into a taxonomy with **17 categories**.

Taxonomy Overview

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

- **Depth Acquisition and Enhancement Methods**
- **Depth-Guided Manipulation Frameworks**
- **Benchmarks, Datasets, and Evaluation Frameworks**
- **Simulation-to-Real Transfer for Depth-Based Manipulation**

Complete Taxonomy Tree

- Depth Enhancement for Robotic Manipulation Using Camera Depth Models Survey Taxonomy
- Depth Acquisition and Enhancement Methods
  - Monocular Depth Estimation for Manipulation
  - Learning-Based Monocular Depth Prediction (4 papers)
    - [1] Attention-based grasp detection with monocular depth estimation (Phan Xuan Tan, 2024) View paper
    - [5] KineDepth: Utilizing Robot Kinematics for Online Metric Depth Estimation (Zhi, 2024) View paper
    - [21] Rgb-based category-level object pose estimation via depth recovery and adaptive refinement (Hui Yang, 2025) View paper
    - [38] Prompting Depth Anything for 4K Resolution Accurate Metric Depth Estimation (Haotong Lin, 2025) View paper
  - Monocular Depth for Specific Manipulation Contexts (4 papers)
    - [42] Object localization and depth estimation for eye-in-hand manipulator using mono camera (Muslikhin, 2020) View paper
    - [46] A cascaded CNN-based method for monocular vision robotic grasping (Xiaojun Wu, 2022) View paper
    - [47] 2D vision-based with monocular depth estimation for pose estimation (Pitijit Charoenwuttikajorn, 2025) View paper
    - [48] Monocular robust depth estimation vision system for robotic tasks interventions in metallic targets (C. V. Almagro, 2019) View paper
  - Stereo and Multi-View Depth Systems
  - Learned Stereo Matching for Robotic Environments (4 papers)
    - [2] A learned stereo depth system for robotic manipulation in homes (Krishna Shankar, 2022) View paper
    - [10] DRoMa:Disparity Diffusion-based Depth Sensing for Material-Agnostic Robotic Manipulation (S Wei, 2024) View paper
    - [11] ClearDepth: enhanced stereo perception of transparent objects for robotic manipulation (Zeng, 2024) View paper
    - [30] Beyond Trade-Off: An Optimized Binocular Stereo Vision Based Depth Estimation Algorithm for Designing Harvesting Robot in Orchards (Li Zhang, 2023) View paper
  - Multi-View Geometric Reconstruction (2 papers)
    - [13] Obtaining an object's 3D model using dual-arm robotic manipulation and stationary depth sensing (Sho Kobayashi, 2022) View paper
    - [27] Rgbgrasp: Image-based object grasping by capturing multiple views during robot arm movement with neural radiance fields (Chang Liu, 2024) View paper
  - Depth Completion and Refinement for Challenging Materials
  - Transparent Object Depth Completion (10 papers)
    - [3] Transparent object depth perception network for robotic manipulation based on orientation-aware guidance and texture enhancement (Yunhui Yan, 2024) View paper

        - [4] Rethinking Transparent Object Grasping: Depth Completion with Monocular Depth Estimation and Instance Mask (Gao XinKai, 2025) View paper
        - [6] Clear grasp: 3d shape estimation of transparent objects for manipulation (Shreeyak S. Sajjan, 2020) View paper
        - [12] Transcg: A large-scale real-world dataset for transparent object depth completion and a grasping baseline (Fang, 2022) View paper
        - [16] Transdiff: Diffusion-based method for manipulating transparent objects using a single rgb-d image (Zhou, 2025) View paper
        - [22] Transparent Object Depth Completion (Zhou Yi-Fan, 2024) View paper
        - [24] A4T: Hierarchical affordance detection for transparent objects depth reconstruction and manipulation (Jiaqi Jiang, 2022) View paper
        - [29] Asgrasp: Generalizable transparent object reconstruction and 6-dof grasp detection from rgb-d active stereo camera (Jun Shi, 2024) View paper
        - [43] CrysFormer++: Dual-phase Refinement Learning for Transparent Object Depth Estimation (Xiao-Mei Zhang, 2025) View paper
        - [49] DCIRNet: Depth Completion with Iterative Refinement for Dexterous Grasping of Transparent and Reflective Objects (Xie Guanghu, 2025) View paper
    - General Depth Refinement and Enhancement (2 papers)
        - [17] Multi-scale progressive fusion-based depth image completion and enhancement for industrial collaborative robot applications (Chuhua Xian, 2024) View paper
        - [50] ReCAP2: Rectified and Context-Aware Polarization Prompting for Robust Depth Enhancement (Zhen-yu Liu, 2025) View paper
    - Specialized Depth Sensing Modalities (2 papers)
    - [26] Effective Marine Monitoring with Multimodal Sensing and Improved Underwater Robotic Perception towards Environmental Protection and Smart Energy â¦ (H Farhadi Tolie, 2024) View paper
    - [28] Polarimetric Imaging for Robot Perception: A Review (Camille Taglione, 2024) View paper
- Depth-Guided Manipulation Frameworks
    - Grasp Detection and Synthesis Using Depth
    - Point Cloud-Based Grasp Generation (2 papers)
        - [7] 6-DOF GraspNet: Variational Grasp Generation for Object Manipulation (Arsalan Mousavian, 2019) View paper
        - [41] Grasping of unknown objects using deep convolutional neural networks based on depth images (Philipp Schmidt, 2018) View paper
    - RGB-D Fusion for Grasp Detection (4 papers)
        - [20] Collision-free grasp detection from color and depth images (Dinh Cuong Hoang, 2024) View paper
        - [33] Depth Estimation Using Monocular Camera for Real-World Multi-Object Grasp Detection for Robotic Arm (Vayam Jain, 2023) View paper
        - [35] Deep robotic grasping prediction with hierarchical RGB-D fusion (Yaoxian Song, 2022) View paper
        - [45] RGB-D grasp detection via depth guided learning with cross-modal attention (Ran Qin, 2023) View paper
    - Affordance-Based and Hierarchical Grasp Planning (2 papers)
        - [32] World Models for General Surgical Grasping (Lin, 2024) View paper
        - [44] Robotic grasping of novel objects using vision (Ashutosh Saxena, 2008) View paper
    - Pose Estimation and Localization with Depth
    - CAD-Based and Model-Driven Pose Estimation (1 papers)
        - [15] Bayesian inference for CAD-based pose estimation on depth images for robotic manipulation (Redick, 2024) View paper
    - Category-Level and Learning-Based Pose Estimation (2 papers)
        - [37] Deep Learning-based Mobile Robot Target Object Localization and Pose Estimation Research (Caixia He, 2023) View paper
        - [39] RGB and 3D-Segmentation Data Combination for the Autonomous Object Manipulation in Personal Care Robotics (Giovanni Mezzina, 2021) View paper
    - 3D Representation Learning for Manipulation
    - Foundation Model-Based 3D Manipulation (3 papers)
        - [8] Lift3d foundation policy: Lifting 2d large-scale pretrained models for robust 3d robotic manipulation (Yueru Jia, 2024) View paper
        - [14] Lift3D Policy: Lifting 2D Foundation Models for Robust 3D Robotic Manipulation (Yueru Jia, 2025) View paper
        - [19] Spatial RoboGrasp: Generalized Robotic Grasping Control Policy (Huang, 2025) View paper
    - Task-Specific 3D Scene Representations (2 papers)
        - [9] Persistent Object Gaussian Splat (POGS) for Tracking Human and Robot Manipulation of Irregularly Shaped Objects (Yu, 2025) View paper
        - [25] 3d shape perception from monocular vision, touch, and shape priors (Wang, 2018) View paper
    - Integrated Vision Systems for Manipulation (3 papers)
    - [31] Exploring the visual space to improve depth perception in robot teleoperation using augmented reality: The role of distance and target's pose in time, success, and â¦ (S ArÃ©valo Arboleda, 2021) View paper
    - [36] Human-Robot Interaction for Assisted Object Grasping by a Wearable Robotic Object Manipulation Aid for the Blind (Lingqiu Jin, 2020) View paper
    - [40] Vision-based robotic grasping using faster R-CNNâGRCNN dual-layer detection mechanism (Jianguo Duan, 2024) View paper
- Benchmarks, Datasets, and Evaluation Frameworks (3 papers)
    - [18] IAM: Enhancing RGB-D Instance Segmentation with New Benchmarks (Aecheon Jung, 2025) View paper
    - [23] A comprehensive study of 3-D vision-based robot manipulation (Yang Cong, 2021) View paper
    - [34] Visual sensing and depth perception for welding robots and their industrial applications (Ji Wang, 2023) View paper
- Simulation-to-Real Transfer for Depth-Based Manipulation ★ (1 papers)
    - [0] Manipulation as in Simulation: Enabling Accurate Geometry Perception in Robots (Anon et al., 2026) View paper

Narrative

Core task: depth enhancement for robotic manipulation using camera depth models. The field addresses how robots can acquire, refine, and exploit depth information to perform reliable grasping and manipulation in diverse environments. The taxonomy organizes research into four main branches: Depth Acquisition and Enhancement Methods focus on improving raw depth signals through learning-based refinement, stereo reconstruction, and specialized techniques for challenging materials like transparent or reflective objects (e.g., ClearDepth[11], Transparent Object Depth[3]); Depth-Guided Manipulation Frameworks integrate enhanced depth into end-to-end

policies and grasp planning systems (e.g., GraspNet[7], Lift3D Foundation[8]); Benchmarks, Datasets, and Evaluation Frameworks provide standardized testbeds and metrics; and Simulation-to-Real Transfer for Depth-Based Manipulation explores how depth models trained in simulation can generalize to physical systems.

A particularly active line of work targets transparent and reflective objects, where standard depth sensors fail—methods like Clear Grasp[6] and Rethinking Transparent Grasping[4] propose neural completion and physics-informed priors to recover missing geometry. Another contrasting direction emphasizes large-scale foundation models and multi-modal fusion (Lift3D Policy[14], Prompting Depth Anything[38]) that leverage pre-trained representations to generalize across object categories. The original paper, Manipulation as Simulation[0], sits within the Simulation-to-Real Transfer branch and emphasizes bridging the gap between synthetic training environments and real-world deployment. Compared to works like KineDepth[5], which refines depth online during manipulation, or Transparent Depth Completion[22], which focuses on material-specific enhancement, Manipulation as Simulation[0] appears to prioritize robust transfer mechanisms that maintain depth fidelity across the sim-to-real boundary, addressing domain shift challenges that remain central to deploying learned depth models in practice.

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 on simulation-to-real transfer techniques that bridge the gap between synthetic training and real-world depth-based manipulation, emphasizing noise modeling and domain adaptation. The sibling subtopic addresses the infrastructure side of depth-based manipulation research, providing datasets, benchmarks, and evaluation protocols. Both support depth-enhanced robotic manipulation but serve complementary roles: one develops transfer methods while the other establishes evaluation standards.

Similarities: - Both support depth-based robotic manipulation research - Both may involve handling challenging scenarios like transparent objects or noisy depth data - Both contribute to improving real-world deployment of depth-guided manipulation systems

Differences: - Original leaf focuses on algorithmic methods for sim-to-real transfer (noise modeling, domain adaptation), while sibling provides evaluation infrastructure (datasets, benchmarks) - Original leaf emphasizes training strategies and transfer techniques, while sibling emphasizes standardized evaluation and comparative assessment - Original leaf excludes methods without explicit sim-to-real focus, while sibling excludes novel algorithms without dataset contributions - Original leaf targets the domain gap problem, while sibling targets reproducibility and standardized comparison

Suggested Search Directions: - Sim-to-real transfer methods that include benchmark contributions or dataset releases - Evaluation frameworks specifically designed to measure sim-to-real transfer quality in depth-based manipulation - Datasets that explicitly model or capture simulation-to-reality domain gaps for depth sensors

Sibling Subtopics

- **Benchmarks, Datasets, and Evaluation Frameworks** (leaves: 1, papers: 3)
- Scope: Datasets, benchmarks, or comprehensive evaluation studies for depth-based manipulation including transparent objects, RGB-D segmentation, or general manipulation surveys.
- Exclude: Excludes methods proposing novel algorithms without dataset contributions; see Depth Acquisition and Enhancement Methods and Depth-Guided Manipulation Frameworks.

Contributions Analysis

Overall novelty summary. The paper proposes Camera Depth Models (CDMs) as a plugin to enhance depth accuracy from commodity RGB-D sensors for robotic manipulation. It resides in the 'Simulation-to-Real Transfer for Depth-Based Manipulation' leaf, which currently contains only this paper among the 50 surveyed works. This isolation suggests the taxonomy captures a relatively sparse research direction explicitly focused on sim-to-real depth transfer, distinguishing it from the broader 'Depth Acquisition and Enhancement Methods' branch where most depth refinement work clusters. The paper's emphasis on modeling depth camera noise patterns to bridge simulation and reality positions it at the intersection of depth enhancement and domain adaptation.

The taxonomy reveals substantial activity in neighboring areas. The 'Depth Completion and Refinement for Challenging Materials' subtopic contains ten papers addressing transparent objects and general depth enhancement, while 'Grasp Detection and Synthesis Using Depth' includes multiple subtopics with methods fusing RGB-D data for manipulation. The 'Foundation Model-Based 3D Manipulation' leaf explores lifting 2D representations to 3D for generalizable policies. The original paper diverges from these by targeting the upstream problem of depth sensor fidelity rather than downstream task-specific fusion or material-specific completion, though its neural data engine approach shares methodological overlap with learned depth refinement techniques in adjacent leaves.

Among 30 candidates examined, the neural data engine contribution shows the most substantial prior work overlap, with three refutable candidates identified from ten examined. The CDM plugin concept and ByteCameraDepth dataset contributions each examined ten candidates with zero refutations, suggesting these elements may be more distinctive within the limited search scope. The statistics indicate that while the depth noise modeling approach has recognizable precedents in the examined literature, the specific framing as a camera-agnostic plugin and the dataset contribution appear less directly anticipated by the top-30 semantic matches and their citations.

Given the limited search scope of 30 candidates, this assessment captures novelty relative to closely related work but cannot claim exhaustive coverage of depth enhancement or sim-to-real transfer literature. The paper's unique taxonomy position and the dataset's zero refutations suggest potential distinctiveness, though the neural data engine's three refutable candidates indicate this component builds on established noise modeling techniques. A broader search might reveal additional precedents, particularly in computer vision depth estimation or domain randomization literature outside the manipulation-focused scope examined here.

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

Contribution 1: Camera Depth Models (CDMs)

Description: The authors introduce Camera Depth Models, a plug-in solution for depth cameras that processes RGB images and noisy depth signals to produce high-quality, denoised metric depth. CDMs are designed to enhance geometric accuracy for specific depth cameras, enabling robots to perceive 3D information with near-simulation-level precision.

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. Self-supervised depth enhancement

URL: View paper

Brief Assessment

Self-supervised Enhancement[72] focuses on self-supervised depth completion for hole artifacts in real-world depth maps without ground truth, while CDMs are camera-specific models trained on synthesized noise patterns with metric depth supervision. The technical approaches and problem formulations differ fundamentally.

---

2. RGB-guided depth map recovery by two-stage coarse-to-fine dense CRF models

URL: View paper

Brief Assessment

RGB-guided Recovery[74] focuses on optimization-based depth map recovery using dense CRF models for correcting erroneous areas in existing depth maps, not on learning camera-specific neural models that process RGB and raw depth signals to produce metric depth for robotic manipulation.

---

3. PGDENet: Progressive Guided Fusion and Depth Enhancement Network for RGB-D Indoor Scene Parsing

URL: View paper

Brief Assessment

PGDENet[73] focuses on RGB-D scene parsing using depth enhancement for semantic segmentation tasks, not on creating plug-in depth camera models for robotic manipulation with metric depth prediction and sim-to-real transfer.

---

4. Depth map recovery based on a unified depth boundary distortion model

URL: View paper

Brief Assessment

Depth Boundary Distortion[80] focuses on correcting boundary distortions (missing, fake, misaligned boundaries) in depth maps using SSIM-based region identification and weighted median filtering. This is fundamentally different from CDMs, which are camera-specific neural models trained to denoise and produce metric depth from RGB and raw depth inputs using a dual-branch ViT architecture with token fusion.

---

5. Cow depth image restoration method based on RGB guided network with modulation branch in the cowshed environment

URL: View paper

Brief Assessment

Cow Depth Restoration[75] focuses on depth restoration for cow monitoring in agricultural environments using RGB-guided networks with modulation branches. This is a domain-specific application for livestock monitoring, not a general plug-in solution for robotic manipulation with depth cameras as proposed in the original paper's CDMs.

---

6. Real-time shading-based refinement for consumer depth cameras

URL: View paper

Brief Assessment

Real-time Shading Refinement[77] focuses on shape-from-shading techniques for depth refinement using inverse rendering and lighting estimation, operating at 30Hz. This differs fundamentally from CDMs, which use neural networks trained on camera-specific noise patterns to denoise depth via RGB-depth fusion without inverse rendering optimization.

---

7. Diffusiondepth: Diffusion denoising approach for monocular depth estimation

URL: View paper

Brief Assessment

DiffusionDepth[71] focuses on monocular depth estimation using diffusion denoising approaches for single RGB images, not on RGB-guided depth refinement for physical depth cameras with noisy sensor inputs as CDMs do.

---

8. Selfredepth: Self-supervised real-time depth restoration for consumer-grade sensors

URL: View paper

Brief Assessment

SelfReDepth[78] focuses on self-supervised depth restoration for consumer-grade sensors using sequential frames and RGB-guided inpainting, whereas CDMs are designed as plug-in solutions for specific depth cameras to achieve simulation-level accuracy for robotic manipulation tasks. The technical approaches and application domains differ substantially.

---

9. A generic framework for depth reconstruction enhancement

URL: View paper

Brief Assessment

Generic Depth Reconstruction[79] focuses on generic depth refinement for tasks like super resolution, denoising, and deblurring using geometric constraints between depth and normal maps. In contrast, the original paper's CDMs are camera-specific models designed to process RGB images and raw depth signals from particular depth cameras to produce metric depth, trained with camera-specific noise models. The candidate does not address camera-specific noise modeling or metric depth prediction for robotic manipulation.

---

10. Adaptive Depth Enhancement Network for RGB-D Salient Object Detection

URL: View paper

Brief Assessment

Adaptive Depth Enhancement[76] focuses on RGB-D salient object detection for visual segmentation tasks, not robotic manipulation or metric depth prediction for depth cameras. The technical goals and application domains are fundamentally different.

Contribution 2: Neural data engine for depth camera noise modeling

Description: The authors develop a neural data engine that learns and models the noise patterns of depth cameras to synthesize high-quality paired training data in simulation. This includes training hole noise and value noise models on real-world data, then using them to generate realistic noisy depth images for training CDMs.

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. A physics-based noise formation model for extreme low-light raw denoising

URL: View paper

Brief Assessment

Physics-based Noise[64] focuses on modeling noise in extreme low-light raw image sensors (photon shot noise, read noise, quantization noise) for denoising, not depth camera noise patterns for robotic manipulation training data.

---

2. The benefits of depth information for head-mounted gaze estimation

URL: View paper

Brief Assessment

Gaze Estimation Depth[66] focuses on modeling depth sensor noise for gaze estimation in AR/VR headsets, not for robotic manipulation training data generation. The noise modeling serves a different application domain and does not challenge the novelty of the original paper's neural data engine for manipulation tasks.

---

3. Realistic depth image synthesis for 3d hand pose estimation

URL: View paper

Prior Art Analysis

Realistic Depth Synthesis[63] demonstrates prior work on learning depth camera noise patterns for synthetic training data generation. The candidate paper explicitly describes mapping Gaussian variables to specific non-i.i.d. depth noise patterns and transforming noise-free synthetic depth images to realistic-looking images that mimic characteristic noise patterns of specific depth cameras. This directly addresses the same problem of modeling depth sensor noise for generating realistic synthetic training data, predating the original paper's neural data engine approach.

Evidence

Evidence 1 - **Rationale**: Both papers describe learning and modeling depth camera noise patterns to generate realistic synthetic training data. The candidate explicitly maps noise distributions to camera-specific patterns and transforms clean synthetic images to realistic ones, which is the core concept of the original's neural data engine. - **Original**: we develop a neural data engine that generates high-quality paired data from simulation by modeling a depth camera's noise pattern - **Candidate**: this observation motivates us to propose an alternative approach, where hand pose model is primarily trained with synthesized hand depth images that closely mimicking the characteristic noise patterns of a specific depth camera make under consideration. it is achieved by firstly mapping a gaussian d...

Evidence 2 - **Rationale**: Both papers emphasize camera-specific noise modeling. The candidate demonstrates generating camera-specific realistic depth images, showing prior work on tailoring noise patterns to specific depth camera models. - **Original**: To this end, we collect typical depth patterns and construct a dataset for various depth cameras that are commonly used in daily robot experiments. specifically, our dataset spans 10 depth modes from 7 different depth cameras, including different stereo and lidar cameras. - **Candidate**: our approach is capable of generating camera-specific realistic-looking hand depth images with precise annotations

Evidence 3 - **Rationale**: Both approaches train noise models from real data and use them to synthesize realistic noisy depth images from clean synthetic data. The candidate's method of mapping distributions to noise patterns parallels the original's training of noise models for data synthesis. - **Original**: we train two noise models on our collected depth dataset for each camera, which are then used for generating stylized low-quality depth images on open datasets to train cdms - **Candidate**: it is achieved by firstly mapping a gaussian distributed variable to certain specific non-i.i.d. (independent and identically distributed) depth noise pattern, and then transforming a âvanillaâ noise-free synthetic depth image to a realistic-looking image

---

4. DEPTHOR: Depth Enhancement from a Practical Light-Weight dToF Sensor and RGB Image

URL: View paper

Brief Assessment

DEPTHOR[67] focuses on simulating dToF sensor noise from synthetic datasets for depth completion tasks, while the original paper develops noise models (hole and value noise) specifically for various depth camera types (stereo, lidar) to enable sim-to-real manipulation. The technical approaches and application domains differ substantially.

---

5. Synthetic training data in AI-driven quality inspection: The significance of camera, lighting, and noise parameters

URL: View paper

Brief Assessment

Synthetic Training Data[69] focuses on rendering parameters (camera position, lighting, computational noise) for 2D RGB image generation in industrial quality inspection, not depth camera noise modeling or depth image synthesis for robotic manipulation.

---

6. Enhancement of 3D Camera Synthetic Training Data with Noise Models

URL: View paper

Prior Art Analysis

3D Camera Noise[65] demonstrates prior work on modeling depth camera noise patterns for synthetic training data generation. Both papers develop noise models from real-world depth camera data and apply them to synthetic data for training neural networks. The candidate paper specifically models lateral and axial noise from 3D cameras and evaluates the impact of noise levels on network performance, showing that appropriate noise modeling improves generalization to real data. This establishes that the concept of learning noise patterns from real depth sensors and applying them to synthetic training data was explored before the original paper.

Evidence

Evidence 1 - **Rationale**: Both papers describe modeling noise from real depth cameras and applying it to synthetic training data. The candidate explicitly states they model noise from real 3D cameras and apply it to synthetic data, which is the core concept of the original's 'neural data engine'. - **Original**: we develop a neural data engine that generates high-quality paired data from simulation by modeling a depth camera's noise pattern. - **Candidate**: the goal of this paper is to assess the impact of noise in 3d camera-captured data by modeling the noise of the imaging process and applying it on synthetic training data. we compiled a dataset of specifically constructed scenes to obtain a noise model.

Evidence 2 - **Rationale**: Both papers train noise models on collected real-world data and then use these models to generate noisy synthetic training data. The candidate demonstrates this workflow was established prior to the original paper. - **Original**: we train two noise models on our collected depth dataset for each camera, which are then used for generating stylized low-quality depth images on open datasets to train cdms. - **Candidate**: the estimated models can be used to emulate noise in synthetic training data. the added benefit of adding artificial noise is evaluated in an experiment with rendered data for object segmentation.

Evidence 3 - **Rationale**: Both papers collect real-world datasets from depth cameras to learn noise characteristics. The candidate's approach of collecting data to model specific noise types (lateral and axial) parallels the original's multi-camera data collection for noise modeling. - **Original**: to train such models, we developed a multi-camera mount and collected a dataset of rgb-depth pairs from seven cameras across ten depth modes. - **Candidate**: we compiled a dataset of specifically constructed scenes to obtain a noise model. we specifically model lateral noise, affecting the position of captured points in the image plane, and axial noise, affecting the position along the axis perpendicular to the image plane.

---

7. Understanding real world indoor scenes with synthetic data

URL: View paper

Brief Assessment

Synthetic Indoor Scenes[62] focuses on generating synthetic training data from 3D CAD scenes for semantic segmentation, not on modeling depth camera noise patterns. Their noise model is based on existing simulated Kinect noise from prior work rather than learning noise patterns from real depth cameras.

---

8. PatchRefiner: Leveraging Synthetic Data for Real-Domain High-Resolution Monocular Metric Depth Estimation

URL: View paper

Brief Assessment

PatchRefiner[68] focuses on high-resolution monocular metric depth estimation using synthetic data but does not describe modeling depth camera noise patterns or developing noise models for training data generation as the original paper does.

---

9. Improved sensor model for realistic synthetic data generation

URL: View paper

Brief Assessment

Improved Sensor Model[61] focuses on modeling RGB camera sensor artifacts (lens effects, sensor noise) for synthetic image generation to improve semantic segmentation. The original paper models depth camera noise patterns (hole noise, value noise) for generating realistic depth training data. These are fundamentally different sensor modalities and noise characteristics.

---

10. Multimodal deep learning for robust RGB-D object recognition

URL: View paper

Prior Art Analysis

Multimodal Deep Learning[70] demonstrates prior work on modeling depth sensor noise patterns for training data generation. The candidate paper explicitly describes 'a data augmentation scheme for robust learning with depth images by corrupting them with realistic noise patterns,' which directly addresses the problem of modeling depth camera noise for synthetic training data. Both papers tackle the challenge of learning from imperfect depth sensor data by modeling and synthesizing realistic noise patterns, though they apply this to different downstream tasks (object recognition vs. manipulation).

Evidence

Evidence 1 - **Rationale**: Both papers describe modeling depth camera noise patterns to generate training data. The candidate explicitly mentions corrupting depth images with realistic noise patterns, which is conceptually similar to the original's neural data engine that models noise patterns. - **Original**: we develop a neural data engine that generates high-quality paired data from simulation by modeling a depth camera's noise pattern - **Candidate**: a data augmentation scheme for robust learning with depth images by corrupting them with realistic noise patterns

Evidence 2 - **Rationale**: Both papers address the problem of learning from imperfect depth sensor data in robotics contexts, establishing that handling noisy depth data was a recognized challenge before the original paper. - **Original**: To this end, we collect typical depth patterns and construct a dataset for various depth cameras that are commonly used in daily robot experiments. specifically, our dataset spans 10 depth modes from 7 different depth cameras, including different stereo and lidar cameras. - **Candidate**: We focus on learning with imperfect sensor data, a typical problem in real-world robotics tasks.

Evidence 3 - **Rationale**: The candidate's data augmentation scheme that corrupts depth images with realistic noise patterns serves a similar purpose to the original's noise models that generate stylized low-quality depth images for training. - **Original**: we train two noise models on our collected depth dataset for each camera, which are then used for generating stylized low-quality depth images on open datasets to train cdms - **Candidate**: a data augmentation scheme for robust learning with depth images by corrupting them with realistic noise patterns

---

Contribution 3: ByteCameraDepth dataset

Description: The authors collect and release ByteCameraDepth, a multi-camera depth dataset containing over 170,000 RGB-depth pairs from seven different depth cameras across ten depth modes. This dataset captures typical depth patterns and noise characteristics from commonly used depth cameras in robotic experiments.

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. Deep denoising for multiview depth cameras

URL: View paper

Brief Assessment

Deep Denoising[56] focuses on noise removal methods using neural networks for multicamera depth systems, but does not describe collecting or releasing a multi-camera depth dataset with RGB-depth pairs and noise characterization as a contribution.

---

2. Self-supervised deep depth denoising

URL: View paper

Brief Assessment

Self-supervised Denoising[57] collects RGB-D data for training depth denoising models but does not focus on multi-camera depth datasets with systematic noise characterization across different camera types as a primary contribution. Their dataset serves their self-supervised training methodology rather than being released as a standalone multi-camera benchmark resource.

---

3. Unsupervised depth completion and denoising for rgb-d sensors

URL: View paper

Brief Assessment

Unsupervised Depth Completion[53] focuses on unsupervised depth completion and denoising methods without collecting ground truth datasets. It does not present a multi-camera depth dataset with RGB-depth pairs and noise characterization like ByteCameraDepth.

---

4. Animal Pose Tracking: 3D Multimodal Dataset and Token-based Pose Optimization

URL: View paper

Brief Assessment

Animal Pose Tracking[60] focuses on animal pose estimation with multimodal videos (RGB, depth, thermal) for behavioral analysis, not on depth camera noise characterization or RGB-depth pair datasets for robotic manipulation.

---

5. SoundLoc3D: Invisible 3D Sound Source Localization and Classification Using a Multimodal RGB-D Acoustic Camera

URL: View paper

Brief Assessment

SoundLoc3D[51] focuses on audio-visual 3D sound source localization using RGB-D cameras and microphone arrays, not on collecting multi-camera depth datasets with noise characterization for robotic manipulation.

---

6. Usage of RGB-D Multi-Sensor Imaging System for Medical Applications

URL: View paper

Brief Assessment

RGB-D Medical[58] focuses on medical applications (obstructive sleep apnea diagnosis) using multi-view depth scanning for craniofacial imaging, not on creating a general-purpose multi-camera depth dataset with noise characterization for robotic manipulation.

---

7. Multi-camera vision-based synchronous positioning and mapping for green construction of electric substations

URL: View paper

Brief Assessment

Multi-camera Vision[55] focuses on RGB camera collaboration for SLAM in electric substation construction and CO2 tracking, not on depth camera noise characterization or RGB-depth pair datasets for robotic manipulation.

---

8. A Multi-spectral Dataset for Evaluating Motion Estimation Systems

URL: View paper

Brief Assessment

Multi-spectral Dataset[59] focuses on multi-spectral motion estimation with thermal/visible cameras and IMU for trajectory evaluation, not multi-camera depth datasets with RGB-depth pairs for noise characterization in robotic manipulation.

---

9. Data Fusion of RGB and Depth Data with Image Enhancement

URL: View paper

Brief Assessment

RGB Depth Fusion[52] focuses on data fusion methods for combining RGB and depth images in industrial sorting applications, not on creating multi-camera depth datasets with noise characterization for robotic manipulation.

---

10. The robodepth challenge: Methods and advancements towards robust depth estimation

URL: View paper

Brief Assessment

RoboDepth Challenge[54] focuses on robust depth estimation under out-of-distribution scenarios using existing benchmarks (KITTI-C, NYUDepth2-C), not on collecting multi-camera RGB-depth pairs with noise characterization from physical depth cameras.

Appendix: Text Similarity Detection

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

References

- [0] Manipulation as in Simulation: Enabling Accurate Geometry Perception in Robots View paper
- [1] Attention-based grasp detection with monocular depth estimation View paper
- [2] A learned stereo depth system for robotic manipulation in homes View paper
- [3] Transparent object depth perception network for robotic manipulation based on orientation-aware guidance and texture enhancement View paper
- [4] Rethinking Transparent Object Grasping: Depth Completion with Monocular Depth Estimation and Instance Mask View paper
- [5] KineDepth: Utilizing Robot Kinematics for Online Metric Depth Estimation View paper
- [6] Clear grasp: 3d shape estimation of transparent objects for manipulation View paper
- [7] 6-DOF GraspNet: Variational Grasp Generation for Object Manipulation View paper
- [8] Lift3d foundation policy: Lifting 2d large-scale pretrained models for robust 3d robotic manipulation View paper
- [9] Persistent Object Gaussian Splat (POGS) for Tracking Human and Robot Manipulation of Irregularly Shaped Objects View paper
- [10] DRoMa:Disparity Diffusion-based Depth Sensing for Material-Agnostic Robotic Manipulation View paper
- [11] ClearDepth: enhanced stereo perception of transparent objects for robotic manipulation View paper
- [12] Transcg: A large-scale real-world dataset for transparent object depth completion and a grasping baseline View paper
- [13] Obtaining an object's 3D model using dual-arm robotic manipulation and stationary depth sensing View paper
- [14] Lift3D Policy: Lifting 2D Foundation Models for Robust 3D Robotic Manipulation View paper
- [15] Bayesian inference for CAD-based pose estimation on depth images for robotic manipulation View paper
- [16] Transdiff: Diffusion-based method for manipulating transparent objects using a single rgb-d image View paper
- [17] Multi-scale progressive fusion-based depth image completion and enhancement for industrial collaborative robot applications View paper
- [18] IAM: Enhancing RGB-D Instance Segmentation with New Benchmarks View paper
- [19] Spatial RoboGrasp: Generalized Robotic Grasping Control Policy View paper
- [20] Collision-free grasp detection from color and depth images View paper
- [21] Rgb-based category-level object pose estimation via depth recovery and adaptive refinement View paper
- [22] Transparent Object Depth Completion View paper
- [23] A comprehensive study of 3-D vision-based robot manipulation View paper
- [24] A4T: Hierarchical affordance detection for transparent objects depth reconstruction and manipulation View paper
- [25] 3d shape perception from monocular vision, touch, and shape priors View paper
- [26] Effective Marine Monitoring with Multimodal Sensing and Improved Underwater Robotic Perception towards Environmental Protection and Smart Energy â View paper
- [27] Rgbgrasp: Image-based object grasping by capturing multiple views during robot arm movement with neural radiance fields View paper
- [28] Polarimetric Imaging for Robot Perception: A Review View paper
- [29] Asgrasp: Generalizable transparent object reconstruction and 6-dof grasp detection from rgb-d active stereo camera View paper

- [30] Beyond Trade-Off: An Optimized Binocular Stereo Vision Based Depth Estimation Algorithm for Designing Harvesting Robot in Orchards View paper
- [31] Exploring the visual space to improve depth perception in robot teleoperation using augmented reality: The role of distance and target's pose in time, success, and â View paper
- [32] World Models for General Surgical Grasping View paper
- [33] Depth Estimation Using Monocular Camera for Real-World Multi-Object Grasp Detection for Robotic Arm View paper
- [34] Visual sensing and depth perception for welding robots and their industrial applications View paper
- [35] Deep robotic grasping prediction with hierarchical RGB-D fusion View paper
- [36] Human-Robot Interaction for Assisted Object Grasping by a Wearable Robotic Object Manipulation Aid for the Blind View paper
- [37] Deep Learning-based Mobile Robot Target Object Localization and Pose Estimation Research View paper
- [38] Prompting Depth Anything for 4K Resolution Accurate Metric Depth Estimation View paper
- [39] RGB and 3D-Segmentation Data Combination for the Autonomous Object Manipulation in Personal Care Robotics View paper
- [40] Vision-based robotic grasping using faster R-CNNâGRCNN dual-layer detection mechanism View paper
- [41] Grasping of unknown objects using deep convolutional neural networks based on depth images View paper
- [42] Object localization and depth estimation for eye-in-hand manipulator using mono camera View paper
- [43] CrysFormer++: Dual-phase Refinement Learning for Transparent Object Depth Estimation View paper
- [44] Robotic grasping of novel objects using vision View paper
- [45] RGB-D grasp detection via depth guided learning with cross-modal attention View paper
- [46] A cascaded CNN-based method for monocular vision robotic grasping View paper
- [47] 2D vision-based with monocular depth estimation for pose estimation View paper
- [48] Monocular robust depth estimation vision system for robotic tasks interventions in metallic targets View paper
- [49] DCIRNet: Depth Completion with Iterative Refinement for Dexterous Grasping of Transparent and Reflective Objects View paper
- [50] ReCAP2: Rectified and Context-Aware Polarization Prompting for Robust Depth Enhancement View paper
- [51] SoundLoc3D: Invisible 3D Sound Source Localization and Classification Using a Multimodal RGB-D Acoustic Camera View paper
- [52] Data Fusion of RGB and Depth Data with Image Enhancement View paper
- [53] Unsupervised depth completion and denoising for rgb-d sensors View paper
- [54] The robodepth challenge: Methods and advancements towards robust depth estimation View paper
- [55] Multi-camera vision-based synchronous positioning and mapping for green construction of electric substations View paper
- [56] Deep denoising for multiview depth cameras View paper
- [57] Self-supervised deep depth denoising View paper
- [58] Usage of RGB-D Multi-Sensor Imaging System for Medical Applications View paper
- [59] A Multi-spectral Dataset for Evaluating Motion Estimation Systems View paper
- [60] Animal Pose Tracking: 3D Multimodal Dataset and Token-based Pose Optimization View paper
- [61] Improved sensor model for realistic synthetic data generation View paper
- [62] Understanding real world indoor scenes with synthetic data View paper
- [63] Realistic depth image synthesis for 3d hand pose estimation View paper
- [64] A physics-based noise formation model for extreme low-light raw denoising View paper
- [65] Enhancement of 3D Camera Synthetic Training Data with Noise Models View paper
- [66] The benefits of depth information for head-mounted gaze estimation View paper
- [67] DEPTHOR: Depth Enhancement from a Practical Light-Weight dToF Sensor and RGB Image View paper
- [68] PatchRefiner: Leveraging Synthetic Data for Real-Domain High-Resolution Monocular Metric Depth Estimation View paper
- [69] Synthetic training data in AI-driven quality inspection: The significance of camera, lighting, and noise parameters View paper
- [70] Multimodal deep learning for robust RGB-D object recognition View paper
- [71] Diffusiondepth: Diffusion denoising approach for monocular depth estimation View paper
- [72] Self-supervised depth enhancement View paper
- [73] PGDENet: Progressive Guided Fusion and Depth Enhancement Network for RGB-D Indoor Scene Parsing View paper
- [74] RGB-guided depth map recovery by two-stage coarse-to-fine dense CRF models View paper
- [75] Cow depth image restoration method based on RGB guided network with modulation branch in the cowshed environment View paper
- [76] Adaptive Depth Enhancement Network for RGB-D Salient Object Detection View paper
- [77] Real-time shading-based refinement for consumer depth cameras View paper
- [78] Selfredepth: Self-supervised real-time depth restoration for consumer-grade sensors View paper
- [79] A generic framework for depth reconstruction enhancement View paper
- [80] Depth map recovery based on a unified depth boundary distortion model View paper