Novelty Assessment Report

Paper: Monocular Normal Estimation via Shading Sequence Estimation

PDF URL: https://openreview.net/pdf?id=d7itDxMD1n

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2026-01-05

Abstract

Monocular normal estimation aims to estimate normal map from a single RGB image of an object under arbitrary lighting. Existing methods rely on deep models to directly predict normal maps. However, they often suffer from 3D misalignment: while the estimated normal maps may appear to have an overall correct color distribution, the reconstructed surfaces frequently fail to align with the geometry details. We argue that this misalignment stems from the current paradigm: the model struggles to distinguish and reconstruct spatially-various geometric, as they are represented in normal maps only by relatively subtle color variations. To address this issue, we propose a new paradigm that reformulates normal estimation as shading sequence estimation, where shading sequences are more sensitive to various geometry information. Building on this paradigm, we present RoSE, a method that leverages image-to-video generative models to predict shading sequences. The predicted shading sequences are then converted into normal maps by solving a simple ordinary least-squares problem. To enhance robustness and better handle complex objects, RoSE is trained on a synthetic dataset, dataset, with diverse shapes, materials, and light conditions. Experiments demonstrate that RoSE achieves state-of-the-art performance on real-world benchmark datasets for object-based monocular normal estimation. Codes and dataset will be released to facilitate reproducible research.

Core Task Landscape

This paper addresses: **monocular normal estimation from single RGB images**

A total of **44 papers** were analyzed and organized into a taxonomy with **19 categories**.

Taxonomy Overview

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

- **Joint Depth and Normal Estimation**
- **Generative and Foundation Models for Geometry Estimation**
- **Structured Scene Understanding for Normal Estimation**
- **Data-Driven Deep Learning Approaches**
- **Specialized Application Domains**
- **Multi-Modal and Sensor Fusion Methods**
- **Novel Paradigms and Representations**
- **3D Reconstruction and Scene Understanding Integration**

Complete Taxonomy Tree

- monocular normal estimation from single RGB images Survey Taxonomy
- Joint Depth and Normal Estimation
  - Geometric Consistency-Based Joint Estimation (4 papers)
  - [1] Adaptive surface normal constraint for geometric estimation from monocular images (Long, 2024) View paper
  - [16] Geonet: Geometric neural network for joint depth and surface normal estimation (Xiaojuan Qi, 2018) View paper
  - [23] IronDepth: Iterative Refinement of Single-View Depth using Surface Normal and its Uncertainty (Bae, 2022) View paper
  - [31] Self-supervised learning for single view depth and surface normal estimation (Huangying Zhan, 2019) View paper
  - Normal-Distance Parameterization Methods (2 papers)
  - [12] NDDepth: Normal-Distance Assisted Monocular Depth Estimation and Completion (Shuwei Shao, 2024) View paper
  - [14] NDDepth: Normal-Distance Assisted Monocular Depth Estimation (Shuwei Shao, 2023) View paper
  - Multi-Task Architectures with Information Fusion (3 papers)
  - [17] Multi-stage information diffusion for joint depth and surface normal estimation (Zhiheng Fu, 2023) View paper
  - [28] Joint prediction of depths, normals and surface curvature from rgb images using cnns (Thanuja Dharmasiri, 2017) View paper
  - [40] Prediction of Depths, Normals, and Surface Curvature from RGB Images using CNNs (Thushara, 2021) View paper
- Generative and Foundation Models for Geometry Estimation
  - Diffusion-Based Geometry Generation (4 papers)
  - [4] GeoWizard: Unleashing the Diffusion Priors for 3D Geometry Estimation from a Single Image (Fu Xiao, 2024) View paper
  - [5] Wonder3D: Single Image to 3D Using Cross-Domain Diffusion (Xiaoxiao Long, 2023) View paper
  - [10] Orchid: Image Latent Diffusion for Joint Appearance and Geometry Generation (Krishnan, 2025) View paper
  - [44] Orchid: Image Latent Diffusion for Joint Appearance and Geometry Generation Supplementary Material (Diff, n.d.) View paper
  - Zero-Shot Metric Geometry Foundation Models (1 papers)
  - [3] Metric3D v2: A Versatile Monocular Geometric Foundation Model for Zero-Shot Metric Depth and Surface Normal Estimation (Hu Mu, 2024) View paper
- Structured Scene Understanding for Normal Estimation
  - Planar Surface Regularization (2 papers)

- [18] Surge: Surface regularized geometry estimation from a single image (Peng Wang, 2016) View paper
- [25] Adaptively Joint Pixel-wise Semantic Correlation in Surface Normal Estimation (Jie Wu, 2022) View paper
- Manhattan World and Vanishing Point Constraints (3 papers)
- [11] GroundNet: Monocular ground plane normal estimation with geometric consistency (Yunze Man, 2019) View paper
- [13] U-ARE-ME: Uncertainty-Aware Rotation Estimation in Manhattan Environments (Aalok Patwardhan, 2024) View paper
- [21] VPLNet: Deep Single View Normal Estimation With Vanishing Points and Lines (Rui Wang, 2020) View paper
- Semantic-Guided Normal Prediction (1 papers)
- [19] Deep surface normal estimation on the 2-sphere with confidence guided semantic attention (Quewei Li, 2020) View paper

- Data-Driven Deep Learning Approaches
  - Encoder-Decoder and Multi-Scale Architectures (3 papers)
  - [7] Depth and surface normal estimation from monocular images using regression on deep features and hierarchical crfs (Bo Li, 2015) View paper
  - [15] Multi-stage cascaded deconvolution for depth map and surface normal prediction from single image (Ram Prasad Padhy, 2019) View paper
  - [20] Encoder-Decoder Structure With the Feature Pyramid for Depth Estimation From a Single Image (Mengxia Tang, 2021) View paper
  - Uncertainty-Aware Normal Estimation (1 papers)
  - [41] Estimating and Exploiting the Aleatoric Uncertainty in Surface Normal Estimation (Gwangbin Bae, 2021) View paper
  - Discriminative Training with Geometric Features (2 papers)
  - [34] Discriminatively trained dense surface normal estimation (Bernhard Zeisl, 2014) View paper
  - [43] Designing Deep Networks for Surface Normal Estimation (Wang Xiao-long, 2014) View paper
- Specialized Application Domains
  - Face and Human-Centric Normal Estimation (1 papers)
  - [22] Cross-modal deep face normals with deactivable skip connections (Victoria Abrevaya, 2020) View paper
  - Omnidirectional and 360° Normal Estimation (2 papers)
  - [35] 360Â° Surface Regression with a Hyper-Sphere Loss (Antonios Karakottas, 2019) View paper
  - [37] 360^o Surface Regression with a Hyper-Sphere Loss (Antonis Karakottas, 2019) View paper
  - Robotics and Navigation Applications (3 papers)
  - [24] Structure-slam: Low-drift monocular slam in indoor environments (Yanyan Li, 2020) View paper
  - [32] Efficient Mobile Robot Navigation Using Quantized Monocular Surface Normal Estimation (Abdullah Man, 2025) View paper
  - [33] Beyond RGB and Events: Enhancing Object Detection under Adverse Lighting with Monocular Normal Maps (Liu Mingjie, 2025) View paper
  - Domain-Specific Normal Estimation (3 papers)
  - [2] Polarimetric monocular leaf normal estimation model for plant phenotyping (Fuduo Xue, 2023) View paper
  - [30] Reconstruction and Understanding of Road Surface Scenes Using Dashboard RGB Cameras (å¼µæ¶µç, 2025) View paper
  - [39] Indoor Image Surface Normal Estimation Using Convolutional Neural Network (ì´ìì¤, 2018) View paper
- Multi-Modal and Sensor Fusion Methods (1 papers)
  - [6] DeepLiDAR: Deep Surface Normal Guided Depth Prediction for Outdoor Scene From Sparse LiDAR Data and Single Color Image (Jiaxiong Qiu, 2018) View paper
- Novel Paradigms and Representations
  - Shading-Based Normal Estimation ★ (1 papers)
  - [0] Monocular Normal Estimation via Shading Sequence Estimation (Anon et al., 2026) View paper
  - Canonical Frame and Local Coordinate Estimation (1 papers)
  - [9] FrameNet: Learning Local Canonical Frames of 3D Surfaces from a Single RGB Image (Jingwei Huang, 2022) View paper
- 3D Reconstruction and Scene Understanding Integration (7 papers)
  - [8] Neuris: Neural reconstruction of indoor scenes using normal priors (Jiepeng Wang, 2022) View paper
  - [26] Dense monocular reconstruction using surface normals (Chamara Saroj Weerasekera, 2017) View paper
  - [27] Research on depth estimation and human body reconstruction effects in single-image scene reconstruction (Xiuyuan Zhou, 2025) View paper
  - [29] StereoSpace: Depth-Free Synthesis of Stereo Geometry via End-to-End Diffusion in a Canonical Space (Tjark Behrens, 2025) View paper
  - [36] MonoClothCap: Towards Temporally Coherent Clothing Capture from Monocular RGB Video (Donglai Xiang, 2020) View paper
  - [38] Superb Monocular Depth Estimation Based on Transfer Learning and Surface Normal Guidance. (Kang Huang, 2020) View paper
  - [42] Normal Estimation Color Image Sensor Depth = 0 A x Global Optimization Surface Normal Output Depth Boundary Detection (Yinda Zhang, 2018) View paper

Narrative

Core task: monocular normal estimation from single RGB images. The field has evolved into several major branches that reflect different modeling philosophies and application contexts. Joint Depth and Normal Estimation approaches treat geometry prediction as a coupled problem, leveraging shared representations to improve both outputs. Generative and Foundation Models for Geometry Estimation harness large-scale pre-training and diffusion-based architectures to produce robust predictions across diverse scenes, as seen in works like GeoWizard[4] and Metric3D v2[3]. Structured Scene Understanding emphasizes semantic and geometric priors—such as planar constraints or Manhattan-world assumptions—to guide normal prediction in indoor or urban environments. Data-Driven Deep Learning Approaches focus on end-to-end architectures and loss formulations that directly optimize normal accuracy, while Specialized Application Domains target specific use cases like autonomous driving or cloth capture. Multi-Modal and Sensor Fusion Methods integrate additional cues such as LiDAR or polarization data to refine estimates. Novel Paradigms and Representations explore alternative input modalities or intermediate representations, and 3D Reconstruction and Scene Understanding Integration connects normal estimation to broader reconstruction pipelines.

Within Novel Paradigms and Representations, a small cluster of works investigates unconventional input signals or intermediate features. Shading Sequence Normal[0] sits in this branch, emphasizing shading cues as a primary source of geometric information—an approach that contrasts with purely data-driven methods that treat RGB pixels as generic features. This focus on shading-based reasoning aligns with classical photometric techniques but leverages modern learning frameworks. Nearby, Adaptive Surface Normal[1] explores adaptive mechanisms for handling varying surface properties, while Polarimetric Leaf Normal[2] demonstrates how polarization can complement RGB for specific materials. The original paper's emphasis on shading sequences distinguishes it from foundation models like

GeoWizard[4] or Wonder3D[5], which rely on large-scale pre-training rather than explicit physical modeling. This positioning highlights an ongoing tension in the field: whether to exploit domain-specific cues or to scale generic architectures.

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

Both subtopics address geometric surface property estimation from single RGB images, going beyond simple color-to-normal regression. Shading-Based Normal Estimation focuses on exploiting photometric cues (shading patterns) to infer surface orientation, while Canonical Frame and Local Coordinate Estimation aims to recover complete local coordinate systems at each surface point. Both represent specialized geometric reasoning approaches rather than generic data-driven mappings.

Similarities: - Both estimate surface geometry properties from monocular RGB images - Both go beyond standard direct regression by incorporating geometric or photometric structure - Both produce per-pixel geometric information (normals or local frames) - Both exclude generic data-driven color-to-normal regression methods

Differences: - Shading-Based methods exploit photometric cues (illumination, shading sequences) while Canonical Frame methods focus on local coordinate system structure - Shading-Based approaches output surface normals through shading analysis, whereas Canonical Frame methods predict full tangent-normal bases or coordinate frames - Shading-Based methods reformulate the problem around shading prediction/interpretation, while Canonical Frame methods emphasize local geometric reference frame establishment - Canonical Frame methods provide richer geometric information (full local frames) compared to Shading-Based methods (primarily normals)

Suggested Search Directions: - Hybrid methods combining shading cues with local frame prediction - Photometric stereo approaches adapted for single-image normal estimation - Methods predicting both shading and local coordinate frames jointly

Sibling Subtopics

- **Canonical Frame and Local Coordinate Estimation** (leaves: 1, papers: 1)
- Scope: Techniques predicting local canonical coordinate frames or tangent-normal bases at each surface point from RGB images.
- Exclude: Methods predicting only surface normals without full local frames belong elsewhere.

Contributions Analysis

Overall novelty summary. The paper proposes reformulating monocular normal estimation as shading sequence estimation, arguing that shading sequences better capture geometric variations than direct normal prediction. According to the taxonomy, this work resides in the 'Shading-Based Normal Estimation' leaf under 'Novel Paradigms and Representations', where it is currently the only paper. This isolation suggests the shading-sequence paradigm represents a relatively unexplored direction within a field that has primarily focused on direct regression or joint depth-normal estimation. The sparse population of this leaf contrasts with more crowded branches like 'Encoder-Decoder and Multi-Scale Architectures' or 'Diffusion-Based Geometry Generation', indicating the approach occupies a niche position.

The taxonomy reveals that neighboring research directions emphasize different modeling philosophies. The sibling leaf 'Canonical Frame and Local Coordinate Estimation' explores local coordinate systems rather than shading cues, while the broader 'Generative and Foundation Models' branch leverages large-scale pre-training without explicit physical modeling. The 'Data-Driven Deep Learning Approaches' branch, containing multiple encoder-decoder architectures, treats RGB pixels as generic features rather than decomposing them into shading components. The paper's focus on shading sequences bridges classical photometric techniques with modern generative models, positioning it between physics-based reasoning and data-driven learning—a boundary less explored than either extreme.

Among the three contributions analyzed across 18 candidate papers, the shading-sequence paradigm examined 2 candidates with no clear refutations, suggesting limited prior work on this specific formulation. The RoSE method using image-to-video models examined 6 candidates without refutation, indicating the application of video generation to normal estimation may be relatively novel. However, the MultiShade synthetic dataset contribution examined 10 candidates and found 2 refutable instances, suggesting synthetic datasets with diverse materials and lighting are more established in the field. The limited search scope (18 candidates total) means these findings reflect top-K semantic matches rather than exhaustive coverage.

Based on the examined candidates, the shading-sequence reformulation and video-generation approach appear less explored than synthetic dataset creation. The taxonomy structure confirms the paper occupies a sparse research direction, though the small search scope (18 papers) and single-paper leaf status warrant caution. The analysis captures semantic neighbors but cannot rule out relevant work outside the top-K matches or in adjacent computer vision subfields like photometric stereo or intrinsic image decomposition.

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

Contribution 1: New paradigm reformulating normal estimation as shading sequence estimation

Description: The authors propose a paradigm shift where monocular normal estimation is reformulated as predicting a shading sequence under canonical lights, which is more sensitive to geometric variations than directly predicting normal maps. This addresses the 3D misalignment problem in existing methods.

This contribution was assessed against **2 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Recovering facial shape using a statistical model of surface normal direction

URL: View paper

Brief Assessment

Statistical Facial Normal[62] focuses on statistical modeling of surface normal distributions for facial shape recovery using azimuthal equidistant projection, not on reformulating normal estimation as shading sequence prediction under canonical lights.

---

2. Learning single-image 3d reconstruction by generative modelling of shape, pose and shading

URL: View paper

Brief Assessment

Generative Shape Pose[61] focuses on learning 3D mesh reconstruction from single images using shading for shape-from-shading in a generative framework, not on reformulating normal estimation as shading sequence prediction under canonical lights for monocular normal map estimation.

Contribution 2: RoSE method using image-to-video generative model for shading sequence prediction

Description: RoSE leverages image-to-video generative models to predict shading sequences from a single grayscale image, then converts these sequences into normal maps using an ordinary least-squares solver. This approach achieves state-of-the-art performance on benchmark datasets.

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

1. A Shading-Guided Generative Implicit Model for Shape-Accurate 3D-Aware Image Synthesis

URL: View paper

Brief Assessment

Shading Guided Implicit[50] focuses on 3D-aware image synthesis using shading for shape regularization in generative models, not on monocular normal estimation from shading sequences via image-to-video models as in RoSE.

2. Illumination and color in computer generated imagery

URL: View paper

Brief Assessment

Illumination Color CGI[45] discusses illumination models in commercial image generation systems and surface normals for light vectors, but does not address monocular normal estimation, shading sequence prediction, or image-to-video generative models for this purpose.

3. Generative AI for 2.5 D Content Creation with Depth-Guided Object Placement

URL: View paper

Brief Assessment

Depth Guided Placement[48] focuses on 2.5D content creation using depth maps and object placement in Blender, not on normal map estimation or shading sequence prediction using video generative models.

4. Face illumination normalization with shadow consideration

URL: View paper

Brief Assessment

Face Shadow Normalization[49] focuses on face illumination normalization using photometric stereo for face recognition, not on general normal map estimation using image-to-video generative models for shading sequence prediction.

5. Static scene illumination estimation from videos with applications

URL: View paper

Brief Assessment

Static Scene Illumination[47] focuses on estimating environment maps from videos for scene illumination recovery in static scenes, not on monocular normal estimation from single images using shading sequences and video generative models.

6. GeoMan: Temporally Consistent Human Geometry Estimation using Image-to-Video Diffusion

URL: View paper

Brief Assessment

GeoMan[46] focuses on depth and normal estimation for human videos using image-to-video diffusion models, not on shading sequence prediction for general object normal estimation. The technical approaches and problem domains are fundamentally different.

Contribution 3: MultiShade synthetic dataset with diverse materials and lighting

Description: The authors curate MultiShade, a large-scale synthetic dataset built on Objaverse models with material augmentation from MatSynth and diverse lighting conditions (parallel, point, and environment lights). This dataset improves model robustness and generalization to complex real-world scenarios.

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. Normalizing images in various weather and lighting conditions using ColorPix2Pix generative adversarial network

URL: View paper

Brief Assessment

ColorPix2Pix Weather[56] focuses on image normalization for autonomous vehicles under weather and lighting variations, not on normal estimation tasks. The candidate does not address synthetic datasets for normal map estimation or geometric reconstruction.

2. SNE-RoadSeg: Incorporating Surface Normal Information into Semantic Segmentation for Accurate Freespace Detection

URL: View paper

Brief Assessment

SNE RoadSeg[58] focuses on freespace detection for autonomous driving using a synthetic dataset (R2D) for road segmentation, not on normal estimation with diverse materials and lighting conditions as in the original paper's MultiShade dataset.

3. SynthOutdoor: A synthetic dataset for 3D outdoor light estimation

URL: View paper

Brief Assessment

SynthOutdoor[51] focuses on outdoor light direction estimation with solar cycles and distant lighting assumptions, not on diverse material augmentation or multi-light photometric stereo setups for normal estimation.

4. Cross-Domain Synthetic-to-Real In-the-Wild Depth and Normal Estimation for 3D Scene Understanding

URL: View paper

Brief Assessment

Cross Domain Synthetic[57] focuses on omnidirectional outdoor scene understanding with the OmniHorizon dataset, which models urban environments and vegetation with dynamic lighting and scene participants. This differs from MultiShade's object-centric approach with material augmentation from MatSynth for normal estimation tasks.

5. IRS: A Large Synthetic Indoor Robotics Stereo Dataset for Disparity and Surface Normal Estimation

URL: View paper

Brief Assessment

IRS Stereo Dataset[55] focuses on indoor robotics stereo vision with disparity and surface normal ground truth, while MultiShade targets monocular normal estimation with shading sequences under canonical lights. The datasets serve different technical purposes and input modalities.

6. SfPUEL: Shape from Polarization under Unknown Environment Light

URL: View paper

Brief Assessment

Shape from Polarization[60] focuses on polarization-based normal estimation under environment lighting with a synthetic dataset for polarization tasks, while the original paper addresses shading sequence estimation for monocular normal estimation with different rendering objectives and physical measurements.

7. Factorized Inverse Path Tracing for Efficient and Accurate Material-Lighting Estimation

URL: View paper

Brief Assessment

Factorized Inverse Path[54] focuses on inverse rendering for indoor scenes with multi-view HDR observations, not on creating synthetic datasets for monocular normal estimation training.

8. Physically-based rendering for indoor scene understanding using convolutional neural networks

URL: View paper

Prior Art Analysis

Physically Based Rendering[59] demonstrates that prior work exists on creating large-scale synthetic datasets with diverse materials and lighting conditions for indoor scene understanding tasks. The candidate paper introduces a dataset with 500k physically-based rendered images from 45k realistic 3D indoor scenes, incorporating material augmentation and diverse lighting conditions (parallel, point, and environment lights). This predates the original paper's MultiShade dataset and shows similar approaches to material diversity and lighting variation, including the use of MatSynth-like material augmentation and multiple lighting setups.

Evidence

Evidence 1 - **Rationale**: Both papers describe creating large-scale synthetic datasets from 3D models with diverse materials and lighting. The candidate paper's 500k images from 45k scenes predates the original's 90k models, establishing prior work on this scale. - **Original**: we curate a dataset named multishade, featuring diverse shapes, materials, and light conditions to ensure robust generalization. multishade is built upon a list of pre-filtered 3d models (90k) curated from objaverse - **Candidate**: w e introduce a large scale (500k images) synthetic dataset that is created from 45k 3d houses designed by humans [20]. using such realistic indoor 3d environments enable us to create 2d images for training in realistic context settings where support constructs (e.g. such as walls, ceilings, windows...

Evidence 2 - **Rationale**: Both papers emphasize material diversity in their datasets. The candidate paper describes surface materials including reflectance, texture, and transparency for photo-realistic rendering, similar to the original's material augmentation approach. - **Original**: we implement material augmentationto the dataset by either retaining the object's original texture or applying material augmentation. with a probability of 0.5, an additional material is assigned from the matsynth dataset (vecchio & deschaintre, 2024), which contains 5,657 high-quality pbr materials... - **Candidate**: the object models provide surface materials, including reflectance, texture, and transparency , which are used to obtain photo-realistic renderings. one of the important aspects of this dataset is the fact that the indoor layouts, furniture/object alignment, and surface materials are designed by peop...

Evidence 3 - **Rationale**: Both papers describe multiple lighting setups including directional/parallel lights, point/local lights, and environment lighting. The candidate paper's four rendering combinations with various lighting conditions demonstrates prior work on diverse lighting for synthetic datasets. - **Original**: for each object, we render observed images under three lighting setups: (1) parallel lightsrandomly placed around the object; (2) point lightswith randomly sampled positions and intensities; and (3) environment lightsusing high-dynamic-range (hdr) maps - **Candidate**: w e render images from these selected cameras using four combinations of rendering algorithms and lighting conditions, ranging from fast/unrealistic rendering with directional lights using the opengl pipeline to physically-based rendering with local lights using mitsuba.

Evidence 4 - **Rationale**: Both papers describe generating large-scale synthetic datasets with normal maps. The candidate's 500k instances with multiple renders and per-pixel ground truth establishes prior work on this type of comprehensive synthetic dataset generation. - **Original**: all images are rendered using blender at a resolution of 576 x 576 following (v oleti et al., 2024), generating approximately 3 million image-normal pairs. - **Candidate**: w e introduce a dataset with 500k synthetic image instances where each instance consists of three image renders with varying render quality , per-pixel accurate normal map, semantic labels and object boundaries.

9. Neural LightRig: Unlocking Accurate Object Normal and Material Estimation with Multi-Light Diffusion

URL: View paper

Prior Art Analysis

Neural LightRig[53] demonstrates prior work on creating synthetic datasets with diverse materials and lighting for normal estimation. The candidate paper explicitly describes building a synthetic relighting dataset with multi-lighting conditions and material variations, which directly addresses the same problem space as MultiShade. Both datasets aim to improve robustness through diverse lighting and material augmentation, with Neural LightRig[53] predating the original paper's submission.

Evidence

Evidence 1 - **Rationale**: Both papers describe creating synthetic datasets specifically designed for multi-light conditions. Neural LightRig[53] explicitly mentions building a 'synthetic relighting dataset' with multi-lighting, which is the same core concept as MultiShade. - **Original**: we curate a dataset named multishade, featuring diverse shapes, materials, and light conditions to ensure robust generalization. multishade is built upon a list of pre-filtered 3d models (90k) curated from objaverse - **Candidate**: we leverage illumination priors from large-scale diffusion models to build our multi-light diffusion model on a synthetic relighting dataset with dedicated designs

Evidence 2 - **Rationale**: Neural LightRig[53] addresses material estimation alongside geometry, indicating their dataset includes material variations. The paper's focus on 'materials' estimation suggests their training data incorporates diverse material properties. - **Original**: we implement material augmentationto the dataset by either retaining the object's original texture or applying material augmentation. with a probability of 0.5, an additional material is assigned from the matsynth dataset - **Candidate**: recovering the geometry and materials of objects from a single image is challenging due to its under-constrained nature

Evidence 3 - **Rationale**: Both datasets feature diverse lighting conditions with point lights in multiple directions. Neural LightRig[53]'s use of 'point light sources in different directions' parallels MultiShade's multi-lighting setup including point lights. - **Original**: for each object, we render observed images under three lighting setups: (1) parallel lightsrandomly placed around the object; (2) point lightswith randomly sampled positions and intensities; and (3) environment lightsusing high-dynamic-range (hdr) maps - **Candidate**: this diffusion model generates multiple consistent images, each illuminated by point light sources in different directions

10. MERLiN: Single-Shot Material Estimation and Relighting for Photometric Stereo

URL: View paper

Brief Assessment

MERLiN[52] uses a different synthetic dataset from [22] focused on photometric stereo with point and environment lighting, not the MultiShade dataset with parallel lights and ring-light setup described in the original paper.

Appendix: Text Similarity Detection

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

References

- [0] Monocular Normal Estimation via Shading Sequence Estimation View paper
- [1] Adaptive surface normal constraint for geometric estimation from monocular images View paper
- [2] Polarimetric monocular leaf normal estimation model for plant phenotyping View paper
- [3] Metric3D v2: A Versatile Monocular Geometric Foundation Model for Zero-Shot Metric Depth and Surface Normal Estimation View paper
- [4] GeoWizard: Unleashing the Diffusion Priors for 3D Geometry Estimation from a Single Image View paper
- [5] Wonder3D: Single Image to 3D Using Cross-Domain Diffusion View paper
- [6] DeepLiDAR: Deep Surface Normal Guided Depth Prediction for Outdoor Scene From Sparse LiDAR Data and Single Color Image View paper
- [7] Depth and surface normal estimation from monocular images using regression on deep features and hierarchical crfs View paper
- [8] Neuris: Neural reconstruction of indoor scenes using normal priors View paper
- [9] FrameNet: Learning Local Canonical Frames of 3D Surfaces from a Single RGB Image View paper
- [10] Orchid: Image Latent Diffusion for Joint Appearance and Geometry Generation View paper
- [11] GroundNet: Monocular ground plane normal estimation with geometric consistency View paper
- [12] NDDepth: Normal-Distance Assisted Monocular Depth Estimation and Completion View paper
- [13] U-ARE-ME: Uncertainty-Aware Rotation Estimation in Manhattan Environments View paper
- [14] NDDepth: Normal-Distance Assisted Monocular Depth Estimation View paper
- [15] Multi-stage cascaded deconvolution for depth map and surface normal prediction from single image View paper
- [16] Geonet: Geometric neural network for joint depth and surface normal estimation View paper
- [17] Multi-stage information diffusion for joint depth and surface normal estimation View paper
- [18] Surge: Surface regularized geometry estimation from a single image View paper
- [19] Deep surface normal estimation on the 2-sphere with confidence guided semantic attention View paper
- [20] Encoder-Decoder Structure With the Feature Pyramid for Depth Estimation From a Single Image View paper
- [21] VPLNet: Deep Single View Normal Estimation With Vanishing Points and Lines View paper
- [22] Cross-modal deep face normals with deactivable skip connections View paper
- [23] IronDepth: Iterative Refinement of Single-View Depth using Surface Normal and its Uncertainty View paper
- [24] Structure-slam: Low-drift monocular slam in indoor environments View paper
- [25] Adaptively Joint Pixel-wise Semantic Correlation in Surface Normal Estimation View paper
- [26] Dense monocular reconstruction using surface normals View paper
- [27] Research on depth estimation and human body reconstruction effects in single-image scene reconstruction View paper
- [28] Joint prediction of depths, normals and surface curvature from rgb images using cnns View paper
- [29] StereoSpace: Depth-Free Synthesis of Stereo Geometry via End-to-End Diffusion in a Canonical Space View paper
- [30] Reconstruction and Understanding of Road Surface Scenes Using Dashboard RGB Cameras View paper
- [31] Self-supervised learning for single view depth and surface normal estimation View paper
- [32] Efficient Mobile Robot Navigation Using Quantized Monocular Surface Normal Estimation View paper
- [33] Beyond RGB and Events: Enhancing Object Detection under Adverse Lighting with Monocular Normal Maps View paper
- [34] Discriminatively trained dense surface normal estimation View paper
- [35] 360Â° Surface Regression with a Hyper-Sphere Loss View paper
- [36] MonoClothCap: Towards Temporally Coherent Clothing Capture from Monocular RGB Video View paper
- [37] 360^o Surface Regression with a Hyper-Sphere Loss View paper
- [38] Superb Monocular Depth Estimation Based on Transfer Learning and Surface Normal Guidance. View paper
- [39] Indoor Image Surface Normal Estimation Using Convolutional Neural Network View paper
- [40] Prediction of Depths, Normals, and Surface Curvature from RGB Images using CNNs View paper
- [41] Estimating and Exploiting the Aleatoric Uncertainty in Surface Normal Estimation View paper
- [42] Normal Estimation Color Image Sensor Depth = 0 A x Global Optimization Surface Normal Output Depth Boundary Detection View paper
- [43] Designing Deep Networks for Surface Normal Estimation View paper
- [44] Orchid: Image Latent Diffusion for Joint Appearance and Geometry Generation Supplementary Material View paper
- [45] Illumination and color in computer generated imagery View paper
- [46] GeoMan: Temporally Consistent Human Geometry Estimation using Image-to-Video Diffusion View paper
- [47] Static scene illumination estimation from videos with applications View paper
- [48] Generative AI for 2.5 D Content Creation with Depth-Guided Object Placement View paper
- [49] Face illumination normalization with shadow consideration View paper
- [50] A Shading-Guided Generative Implicit Model for Shape-Accurate 3D-Aware Image Synthesis View paper
- [51] SynthOutdoor: A synthetic dataset for 3D outdoor light estimation View paper
- [52] MERLiN: Single-Shot Material Estimation and Relighting for Photometric Stereo View paper
- [53] Neural LightRig: Unlocking Accurate Object Normal and Material Estimation with Multi-Light Diffusion View paper
- [54] Factorized Inverse Path Tracing for Efficient and Accurate Material-Lighting Estimation View paper
- [55] IRS: A Large Synthetic Indoor Robotics Stereo Dataset for Disparity and Surface Normal Estimation View paper
- [56] Normalizing images in various weather and lighting conditions using ColorPix2Pix generative adversarial network View paper
- [57] Cross-Domain Synthetic-to-Real In-the-Wild Depth and Normal Estimation for 3D Scene Understanding View paper

- [58] SNE-RoadSeg: Incorporating Surface Normal Information into Semantic Segmentation for Accurate Freespace Detection View paper
- [59] Physically-based rendering for indoor scene understanding using convolutional neural networks View paper
- [60] SfPUEL: Shape from Polarization under Unknown Environment Light View paper
- [61] Learning single-image 3d reconstruction by generative modelling of shape, pose and shading View paper
- [62] Recovering facial shape using a statistical model of surface normal direction View paper