Intrinsic Image Diffusion for Indoor Single-view Material Estimation

What is the problem?

The problem addressed in the paper is "appearance decomposition" in computer vision, particularly for indoor scenes. This task involves predicting geometric, material, and lighting properties from a single image. The challenge lies in the inherent ambiguity between lighting and material properties—since both interact to create an object's visual appearance, separating these elements from just one image is difficult. Additionally, there is a lack of real-world datasets for training, making it harder to develop accurate models

What has been done earlier?

Earlier work in the field of intrinsic image decomposition began in the 1970s and has included a variety of heuristic-based methods, such as those using image gradients, chromaticity clustering, and depth cues. Later approaches introduced real-world datasets for indoor scenes and used conditional random fields to improve decomposition accuracy. More recent methods have been part of full inverse rendering frameworks, often employing learning-based algorithms trained on increasingly photo-realistic synthetic datasets and utilizing advanced architectures such as UNet, cascaded networks, and Transformers

Raushan Kumar,B421038

Intrinsic Image Diffusion for Indoor Single-view Material Estimation

What are the remaining challenges?

Ambiguity in Decomposition: Existing methods tend to be deterministic and attempt to predict a single solution, which often leads to averaging across possible solutions and losing high-frequency details.

Lack of Real-World Datasets: Most models are trained on synthetic data due to the absence of large-scale real-world intrinsic image decomposition datasets, which introduces a domain gap when applying these models to real-world images

What novel solution proposed by the authors to solve the problem?

The authors propose a probabilistic formulation of the appearance decomposition problem. Instead of predicting a single solution, they employ a conditional generative diffusion model to sample from the solution space. This model leverages the strong prior of recent diffusion models trained on large-scale real-world images. The approach produces significantly sharper, more consistent, and more detailed material estimates. The method also reduces the domain gap by fine-tuning a pre-trained Stable Diffusion model on synthetic data, utilizing the learned priors from real-world images to improve generalization

Raushan Kumar,B421038