From Monocular Depth Estimation to Monocular Metric Depth Estimation

Extending Efficient Depth Perception towards Physically Meaningful 3D Understanding

Md Arafat Islam 2111293642

Motivation

Depth estimation is fundamental for robotics, autonomous systems, AR/VR, and industrial inspection. Traditional sensors are expensive; monocular vision offers scalability.

Monocular Depth Estimation (MDE)

Predicts relative depth from a single RGB image. Depth is scale-ambiguous.

Previous Work Summary

CNN-based U-Net architectures with SE blocks. U-ResNet50 achieved 94.62% SSIM. Quantized and deployed for edge intelligence.

Limitations of MDE

Relative depth cannot provide real-world distance measurements. Not suitable for safety-critical systems.

Why Relative Depth is Insufficient

Industry requires metric distance guarantees, not normalized depth maps.

Monocular Metric Depth Estimation (MMDE)

Predicts absolute depth in meters from a single RGB image. Removes scale ambiguity.

MDE vs MMDE Comparison

MDE: Relative, scale-ambiguous MMDE: Absolute, metric, physically meaningful

Why MMDE is a Natural Extension

Builds on existing architectures. Adds scale recovery and geometric consistency.

Research Scope of MMDE

Scale generalization, camera intrinsics, zero-shot depth, robustness.

Industrial Importance

Autonomous driving, robotics, AR, drones, smart manufacturing.

Investment Perspective

MMDE reduces hardware costs and enables scalable deployment.

Comparative Methods

Supervised, self-supervised, and hybrid MMDE approaches.

State-of-the-Art MMDE

Metric3Dv2, UniDepth, HybridDepth, Depth Anything, ECoDepth.

Research Gap

Lack of efficient, edge-deployable MMDE systems.

Proposed Research Direction

Scale-aware U-ResNet50-based MMDE framework.

Evaluation Strategy

Metric RMSE, scale consistency, cross-dataset testing.

Expected Contributions

Scalable MMDE, industrial relevance, journal-level contribution.

Conclusion

MMDE transforms depth estimation into physically meaningful perception.