الجامعة المصرية اليابانية للعلوم و التكنولوجيا
エジプト日本科学技術大学
EGYPT-JAPAN UNIVERSITY OF SCIENCE AND TECHNOLOGY

CSE 429 Computer Vision and Pattern Recognition

Final Project

Ahmed Mohamed El-Khouly **120210047**

Mohamed Mohy Akl **120210268**

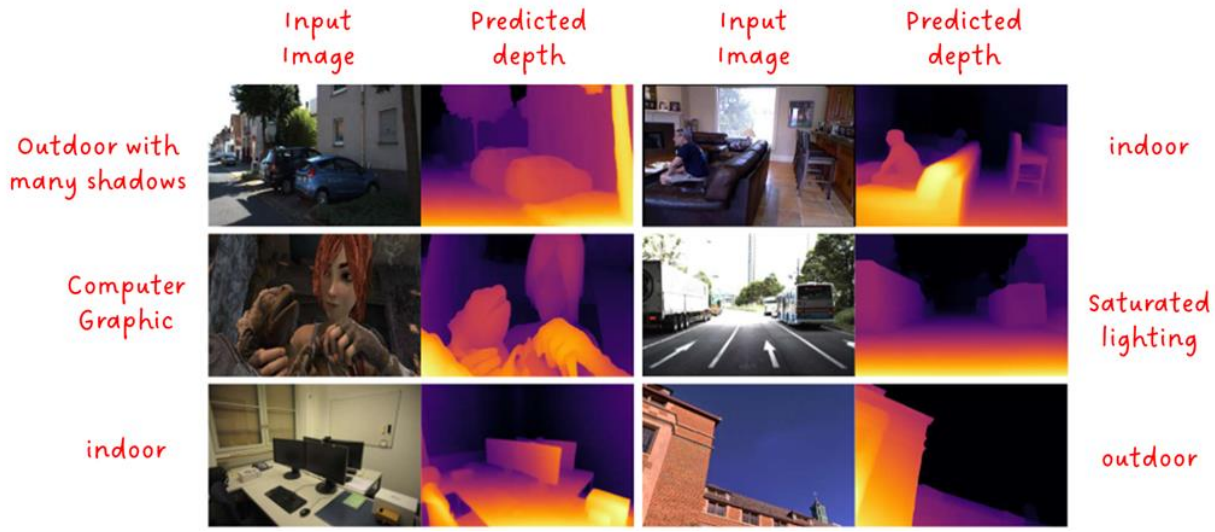
Eyad Amgad Mostafa Mahmoud **120210276**

Submitted to Dr: Ahmed Gomaa

Abstract

This project explores real-time monocular depth estimation using only a standard laptop webcam. We use a pretrained MiDaS model to generate depth maps from single RGB frames. Our system outputs visually realistic depth maps in real-time with a simple and interactive user interface.

Teaser Figure



---

1. Introduction

Depth estimation is a fundamental problem in computer vision with applications in AR, robotics, and accessibility. Traditional methods require stereo vision or depth sensors, but this project tackles it using only monocular input from a laptop camera. We use the MiDaS model, which generalizes well to diverse scenes.

---

2. Approach

- We use the DPT_Large variant of the MiDaS model from PyTorch Hub.
- Input frames from webcam are preprocessed using a transformation pipeline.
- The model outputs a depth map which is then normalized and color-mapped.
- We use Streamlit for the GUI.
- Outputs are optionally saved as a video using OpenCV.

- Challenge: balancing real-time speed with model size. We tested MiDaS Small as an alternative.

---

3. Experiments and Results

- **Setup:** We tested 4 indoor environments, captured via webcam.
- **Dataset:** Real-time input (webcam), and some qualitative tests using NYU Depth v2 images.
- **Metrics:** Visual inspection, FPS, latency, and depth map sharpness.
- **Model Parameters:** Resolution scaling, model size, transform types.
- **Baselines:** Naive method = grayscale conversion for brightness-based depth (not accurate).

Model	FPS	Depth Map Quality
MiDaS Large	6-7	High
MiDaS Small	13+	Medium

We found that MiDaS Small is better for responsiveness, while Large provides clearer depth separation.

---

4. Qualitative Results

Include screenshots of:

- Clear depth edges
- Dark/light variations
- Failure cases (e.g., low light or flat-texture objects)

---

5. Conclusion and Future Work

We demonstrated that monocular depth estimation is feasible in real-time on a laptop using MiDaS. Future work includes:

- FPS optimization
- Adding object detection overlay

- Saving depth maps per frame
- Supporting external webcams or smartphone cameras3

---

6. References

1-Ranftl, René, et al. "Vision Transformers for Dense Prediction." arXiv preprint arXiv:2103.13413 (2021).

2-Liu, Zhaochuan, et al. "MVP-Depth: Multi-view Pretraining for Monocular Depth Estimation." arXiv:2406.19675 [cs.CV]. https://arxiv.org/abs/2406.19675

3-Depth Estimation Using MiDaS – Medium

4-Real-time Depth Estimation on CPU Using MiDaS