Final Project Proposal

” <u>Depth from Single Image</u>”

” <u>Depth Estimation Model</u>”

Mohammad Alshurbaji

CPSC-8810

Deep Learning for Computational Photography

- # The Chosen Topic

For my final project, I have chosen to focus on Depth from a Single Image. This topic has inspired me for various reasons, which will be elaborated upon in the following sections.

- # Problem Definition

Detecting and avoiding barriers is a critical challenge in **robotics**, particularly for autonomous navigation in dynamic and unstructured environments. Traditional robotic systems often rely on expensive and bulky 3D sensors like LiDAR to achieve this functionality, which limits their scalability in cost-sensitive applications. <u>With advancements in Deep Learning, estimating depth from a single image has emerged as a promising alternative to generate 3D environmental data using low-cost monocular cameras.</u>

This project aligns with my background in Mechanical Engineering and my passion for robotics. By implementing a depth estimation model, I aim to enable robots to perceive their surroundings, detect barriers, and avoid collisions effectively.

” <u>**Depth Estimation Model**</u>”

- # Background

Depth estimation has traditionally relied on stereo vision systems or active sensors like LiDAR. Stereo systems require complex calibration between two or more cameras, while active sensors face limitations such as high cost, power consumption, and reduced performance in certain conditions. Recent advancements in convolutional neural networks (CNNs) has revolutionized monocular depth estimation by leveraging large datasets and end-to-end learning. Models like MiDaS and Monodepth2 have demonstrated that depth can be accurately predicted from single images, providing a lightweight and cost-efficient solution for applications such as robotics, autonomous vehicles, and augmented reality.

- # Proposal Method

Deep Learning Method

I will use a convolutional neural network (CNN)-based architecture, such as **MiDaS**, for depth estimation. This model has shown excellent performance in diverse environments and is robust to variations in lighting and texture. Its encoder-decoder structure enables efficient feature extraction and high-resolution depth predictions, making it suitable for real-time robotics applications.

Datasets

- **KITTI Dataset**: A widely used dataset for autonomous driving that provides paired RGB images and depth maps. It has real-world driving data, including urban and semi-rural scenes. While it is one of the most comprehensive datasets available, its large size and the computational limitations of my laptop led me to opt for smaller, more manageable datasets for this project.
- **NYU Depth Dataset V2**: Focused on indoor environments, this dataset includes RGB images and their corresponding depth annotations, ensuring coverage of different scenarios. It has a wide variety of indoor environments and dense depth maps.
- **Make3D (Compact Dataset)**: A lightweight dataset tailored for monocular depth estimation, consisting of outdoor images paired with corresponding depth maps.

Evaluation Metrics

1. **Mean Absolute Error (MAE)**: Measures the average pixel-wise difference between predicted and ground-truth depth.
2. **Root Mean Square Error (RMSE)**: Quantifies the overall error magnitude in depth prediction.