

Monocular Depth Estimation Benchmarking with Thermal Dataset

Authors : Ali Akyar, Osman Serdar Gedik

Abstract : Depth estimation is a challenging computer vision task that involves estimating the distance between objects in a scene and the camera. It predicts how far each pixel in the 2D image is from the capturing point. There are some important Monocular Depth Estimation (MDE) studies that are based on Vision Transformers (ViT). We benchmark three major studies. The first work aims to build a simple and powerful foundation model that deals with any images under any condition. The second work proposes a method by mixing multiple datasets during training and a robust training objective. The third work combines generalization performance and state-of-the-art results on specific datasets. Although there are studies with thermal images too, we wanted to benchmark these three non-thermal, state-of-the-art studies with a hybrid image dataset which is taken by Multi-Spectral Dynamic Imaging (MSX) technology. MSX technology produces detailed thermal images by bringing together the thermal and visual spectrums. Using this technology, our dataset images are not blur and poorly detailed as the normal thermal images. On the other hand, they are not taken at the perfect light conditions as RGB images. We compared three methods under test with our thermal dataset which was not done before. Additionally, we propose an image enhancement deep learning model for thermal data. This model helps extract the features required for monocular depth estimation. The experimental results demonstrate that, after using our proposed model, the performance of these three methods under test increased significantly for thermal image depth prediction.

Keywords : monocular depth estimation, thermal dataset, benchmarking, vision transformers

Conference Title : ICCVIP 2024 : International Conference on Computer Vision and Image Processing

Conference Location : Istanbul, Türkiye

Conference Dates : June 27-28, 2024