Real (True) depth estimation from indoor scenes, given a model (DL tool) for virtual depth estimation

Computer Vision (CS6350)
TPA - 10

1. Problem Statement

Monocular depth estimation models are trained to estimate virtual depth maps, i.e, they are trained to perform ordinal regression and thus they do not estimate real depth (actual distance from camera sensor).

The goal of this TPA is to design a model that can estimate real-world distances for every pixel in the input image from the camera sensor.

2. Input

- Single RGB image (and/or depth map estimated using a DL model).
- Real distances will be provided for some images.

Sample RGB images and the corresponding depth maps:

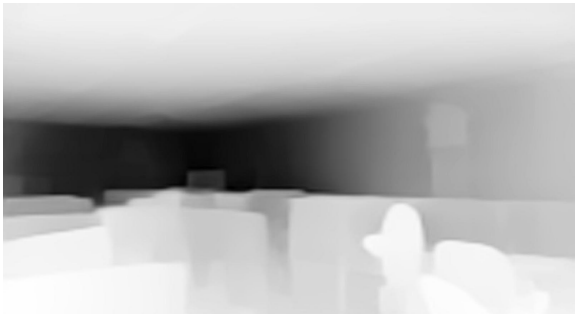
Input Image	Depth Map
	

Figure1: Input Image Virtual depth map

3. Output

Distance (in any units, say meters) for every pixel in the input image.

Sample input images and the corresponding ground truth distance values (recorded using ZED RGBD camera)

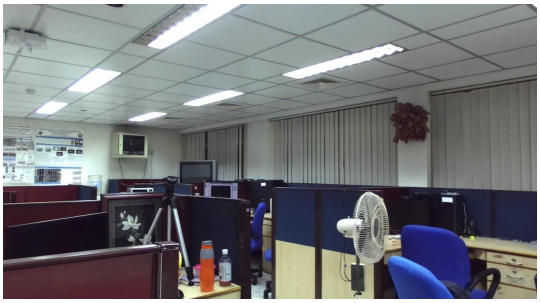
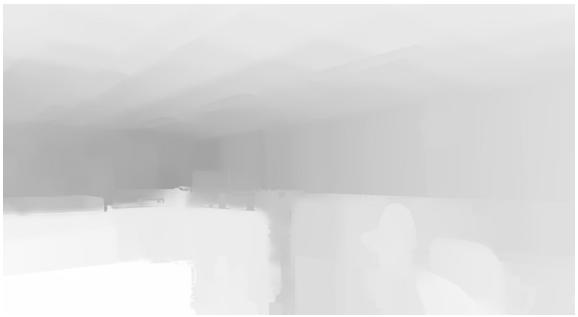
Input Image	Ground truth*
	

Figure 2: Input Image Real depth map

*- Ground truth distance values have been linearly scaled for the purpose of visualization.

4. Dataset

- **NYU-Depth V2:**
  The NYU-Depth V2 data set is comprised of video sequences from a variety of indoor scenes. It features 1449 densely labeled pairs of aligned RGB and depth images.
  Link: https://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html

- **ScanNet:**
  ScanNet is an RGB-D video dataset containing 2.5 million views in more than 1500 scans, annotated with 3D camera poses, surface reconstructions, and instance-level semantic segmentations.
  Link: https://github.com/ScanNet/ScanNet

5. References

[1] Haseeb, Muhammad Abdul, et al. "DisNet: a novel method for distance estimation from monocular camera." 10th Planning, Perception and Navigation for Intelligent Vehicles (PPNIV18), IROS (2018).

[2] Zhu, Jing, and Yi Fang. "Learning object-specific distance from a monocular image." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

[3] Yang, Guanglei, et al. "Transformers solve the limited receptive field for monocular depth prediction." arXiv preprint arXiv:2103.12091 (2021).

[4] Godard, Clément, Oisin Mac Aodha, and Gabriel J. Brostow. "Unsupervised monocular depth estimation with left-right consistency." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[5] Fu, Huan, et al. "Deep ordinal regression network for monocular depth estimation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

Aug, 2022