Predicting LIDAR Intensity from RGB and Depth Images

<TYPE OF THESIS> in computer science

vorgelegt von

Carsten Schmotz

geb. am 23.09.1996 in Dachau

angefertigt am

Department Informatik
Lehrstuhl Graphische Datenverarbeitung
Friedrich-Alexander-Universität Erlangen-Nürnberg

Betreuer: Richard Marcus

Betreuender Hochschullehrer: Prof. Dr. Marc Stamminger

Beginn der Arbeit: 22.07.2024

Abgabe der Arbeit: 29.07.2024

Contents

Abstract

This report explores the application of the Pix2Pix network for predicting LiDAR intensity mpas using RGB images and depth information as additional input. The used dense depth maps are created from the rgb images process thurg Bilateral Propagation Network for Depth Completion and the DepthAnything models (v1 and v2). This appproach used rgb pictures from the kitti dataset and demonstrastes significant imporvements in prediction accuraxy and robustnessover conventional methods

Chapter 1

Introduction

1.1 Motivation

LiDAR senors provide critical depth information for autonomous driving and robotics. The LiDAR intensity maps are often spares and incomplete. Using depht maps as an addtional input is a way to impove the richness and accuracy ot the LiDAR predection. The Pix2Pix network for image-to-image translation offers a promising approch for integrating these modalities.

1.2 Contribution

This project explores the use of the Pix2Pix network to predict LiDAR intensity maps by leveraging RGB images and depth maps as additional inputs.

1.3 Related Work

DeptAnything Models:
DepthAnything represents a significant advancement in monocular depth estimation by leveraging both labeled and unlabeled data at a large scale. Trained on 1.5 million labeled images and over 62 million unlabeled images, DepthAnything achieves state-of-the-art performance in depth estimation tasks. The model excels in both zero-shot relative and metric depth estimation, outperforming previous models such as MiDaS v3.1 and ZoeDepth.
version 2
3D Reconstruction Techniques: 3D reconstruction is a broad field focused on creating three-dimensional models from two-dimensional images or depth data. Techniques in 3D reconstruction include methods like structure-from-motion (SfM), multi-view stereo (MVS), and volumetric approaches. These techniques aim to generate accurate 3D representations of scenes or objects from multiple views or depth sensors.

Chapter 2

Preparations

used google colab, pix2pix network getting the right input, pix2pix problems

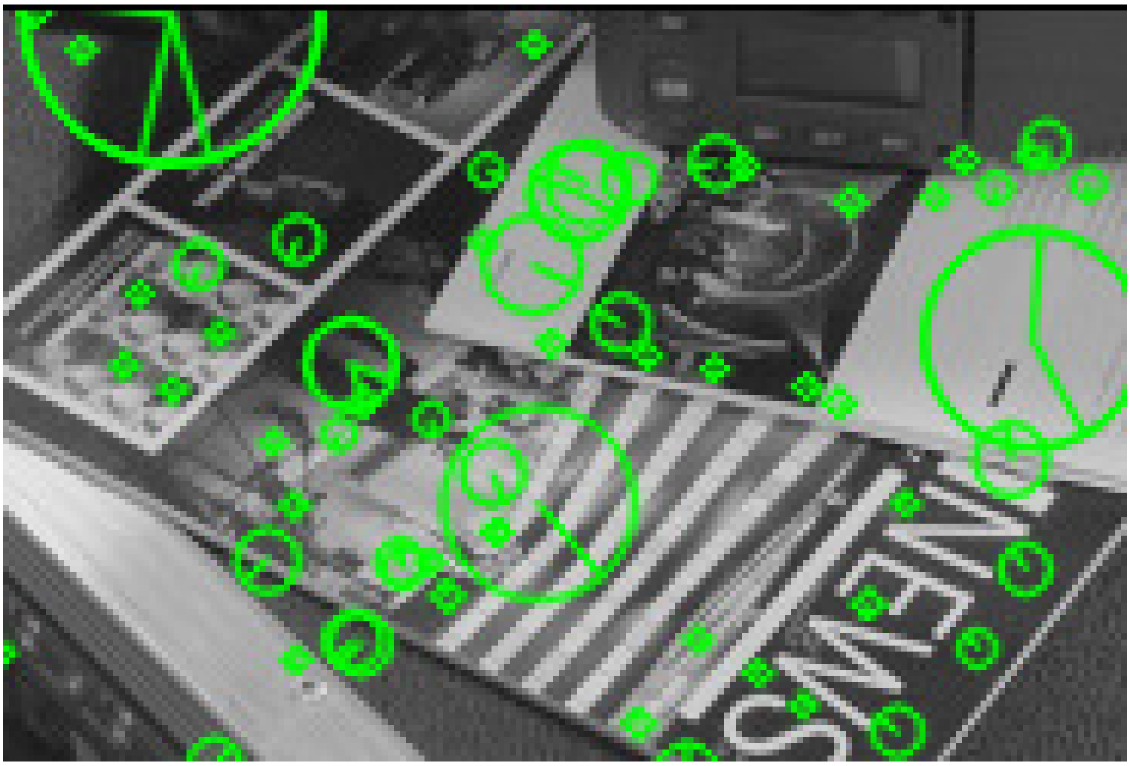

Figure 2.1: caption.

Chapter 3

Predicting LIDAR Intensity from RGB and Depth Images

3.1 Setup

The used the bp net it is for depth completion and depth prediction pix2pix model with modified data loader to get 4 dim. input rgb plus depth used base model for depthanything The setup for this project involved configuring a series of models and datasets to predict LIDAR intensity from RGB and depth images. The key components and their roles are outlined below:

3.1.1 Bilateral Propagation Network (BP Net)

- **Purpose:** Used for depth completion to improve depth maps from incomplete or noisy data.
- **Dataset:** Trained and evaluated on the KITTI dataset to refine depth estimations.

3.1.2 Depth Anything Models

- **Depth Anything v1:** Provided initial depth estimations using large-scale unlabeled data for zero-shot learning.
- **Depth Anything v2:** Enhanced depth prediction capabilities, incorporating improvements over v1 for better depth map accuracy.

3.1.3 Metric Depth Estimation

- **Purpose:** Supplemented depth maps with precise metric depth measurements to enhance model performance.

3.1.4 Dataset

- **Source:** KITTI dataset, containing RGB images and corresponding depth maps.
- **Preparation:** Combined RGB images with depth maps generated by BP Net and Depth Anything models.

3.1.5 pix2pix Network

- **Purpose:** To predict LIDAR intensity from RGB and depth images.
- **Training Data:** RGB and depth pairs from the KITTI dataset.

3.2 Implementation

The implementation details focus on how the above components were practically applied and integrated:

3.2.1 Data Preprocessing

- **Depth Map Generation:** Utilized BP Net and Depth Anything models to generate depth maps from the KITTI dataset. Depth maps were processed to ensure consistency and accuracy before use.
- **Dataset Preparation:** RGB images were paired with the generated depth maps to create a comprehensive training dataset for the pix2pix model.

3.2.2 Model Training

- **pix2pix Configuration:** Adapted the pix2pix architecture to accept RGB images and depth maps as inputs. The network was trained to predict LIDAR intensity values based on these inputs.
- **Training Process:** Configured training parameters such as learning rate, batch size, and number of epochs. The model was trained using the prepared dataset to optimize performance in predicting LIDAR intensity.

3.3 Results

test run rgb only. depht from depthanything the depth from depthanything v2 and metriv form depthanything v2 6 runs with different solution

Chapter 4

Conclusion

4.1 Appendix

ere is a citation for the Depth Anything paper [3], its version 2 [2], and for the paper by Saxena et al. [1]

4.2 References

bp net work pix2pix depth anything v1 2 paper for them some for lidar intensity

List of Figures

Bibliography

[1] A. Saxena, S.H. Chung, and A.Y. Ng. “3-D Depth Reconstruction from a Single Still Image”. In: *International Journal of Computer Vision* 76.1 (2008), pp. 53–69. DOI: 10.1007/s11263-007-0071-y.

[2] Lihe Yang et al. “Depth Anything V2”. In: *arXiv preprint arXiv:2406.09414* (2024).

[3] Lihe Yang et al. “Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2024.