

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

Betreuer Hochschullehrer: Prof. Dr. Marc Stamminger

Beginn der Arbeit: 22.07.2024

Abgabe der Arbeit: 29.07.2024

Contents

1	Introduction	2
1.1	Motivation	2
1.2	Contribution	2
1.3	Related Work	2
2	Preparations	3
3	Predicting LIDAR Intensity	4
3.1	Setup	4
3.1.1	Bilateral Propagation Network (BP Net)	4
3.1.2	Depth Anything Models	4
3.1.3	Pix2pix Network	4
3.2	Implementation	5
3.2.1	Depth Map Generation	5
3.2.2	Model Training	6
3.3	Results	6
3.3.1	Depth BP Net to LIDAR	7
3.3.2	Depth Anything v1 to LIDAR	8
3.3.3	Depth Anything v2 to LIDAR	8
3.3.4	Metric Depth to LIDAR	9
3.3.5	RGB plus BP Net to LIDAR	9
3.3.6	RGB plus Depth Anything v1 to LIDAR	10
3.3.7	RGB plus Depth Anything v2 to LIDAR	10
3.3.8	RGB plus Metric Depth to LIDAR	11
4	Conclusion	15
4.1	Appendix	15
4.1.1	RGB to LIDAR	15
4.2	References	15

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 explores potential improvements in prediction accuracy and robustness over conventional methods.

Chapter 1

Introduction

1.1 Motivation

LiDAR sensors provide critical depth information for autonomous driving and robotics. The LiDAR intensity maps are often sparse and incomplete. Using depth maps as an additional input is a way to improve the richness and accuracy of the LiDAR prediction. The Pix2Pix network for image-to-image translation offers a promising approach 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

DepthAnything 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

Chapter 3

Predicting LIDAR Intensity

3.1 Setup

First of all the depth was created thru rgb pictures taken from "The KITTI Dataset". In order to get a great variatyof depth for comparison Bilateral Propagation Network (BP Net), Depthanything v1 & 2 and metric v2 was used. The metric depth v1 was not used because it did not run intime.

3.1.1 Bilateral Propagation Network (BP Net)

- **Purpose:** Used for depth completion to improve depth maps from incomplete or noisy data.
- **Dataset:** The input were raw data from kitti with a depth completion set already available.

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.
- **Metric Depth Estimation v2:** Supplemented depth maps with precise metric depth measurements.
- **Dataset:** Corresponding RGB images used in BP Net for better comparison.

3.1.3 Pix2pix Network

- **Purpose:** Image to Image processing to predict LIDAR intensity from RGB and depth images.

- **Training Data:** RGB and depth pairs from the KITTI dataset. Eight different input data variations were used see in Chapter 3.2.2.

3.2 Implementation

3.2.1 Depth Map Generation

- **BP Net:** An existing pretrained model on kitti data was loaded. 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. The output was 1000 depthmaps
- **Depthanything v1 & 2 and Metric Depth:** The depthanything networks provide also trained models for kitti. For all depthanything networks there are three versions available. The small, base and large model. For similar results the base model was loaded in each case.
- **Dataset Preparation:** RGB images were paired with the generated depth maps to create a comprehensive training dataset for the pix2pix model.

The three depthmaps sources (shown in picture 3.1) have different properties. The BP Net has very little resolution in the nearfield. The depthanything depthmaps are both more dense in the nearfield but the far distant objects in background are missing. The version 2 has better edge visualizations and some distance objects like the sign on the lefthand side or the car are better visible. The metric depth has the best focus on the farplane from all of them but near field has very little information.



Figure 3.1: RGB and created depth (BP, Deptanything v1,2 and metric v2)

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. used kitti step 1000 pictures, 800 train 100 val 100 test

3.3 Results

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

In this section, we present the results of our experiments using different setups to predict LIDAR intensity. We evaluated the performance of several methods, including RGB to LIDAR, RGB plus depth from BP Net to LIDAR, depth from BP Net to LIDAR, and combinations involving Depth Anything v1 and v2, as well as metric depth estimation. The results from nine runs are summarized below:

3.3.1 Depth BP Net to LIDAR

Using only the depth maps generated by BP Net, without accompanying RGB images, the model was trained to predict LIDAR intensity. This approach yielded a notable improvement over RGB-only predictions, highlighting the value of depth information in enhancing prediction accuracy.



Figure 3.2: RGB LiDAR and predicted LiDAR.

3.3.2 Depth Anything v1 to LIDAR

The Depth Anything v1 model provided initial depth estimations using large-scale unlabeled data. When used to predict LIDAR intensity, this model showed improved results compared to BP Net alone but was less accurate compared to its successor, Depth Anything v2.

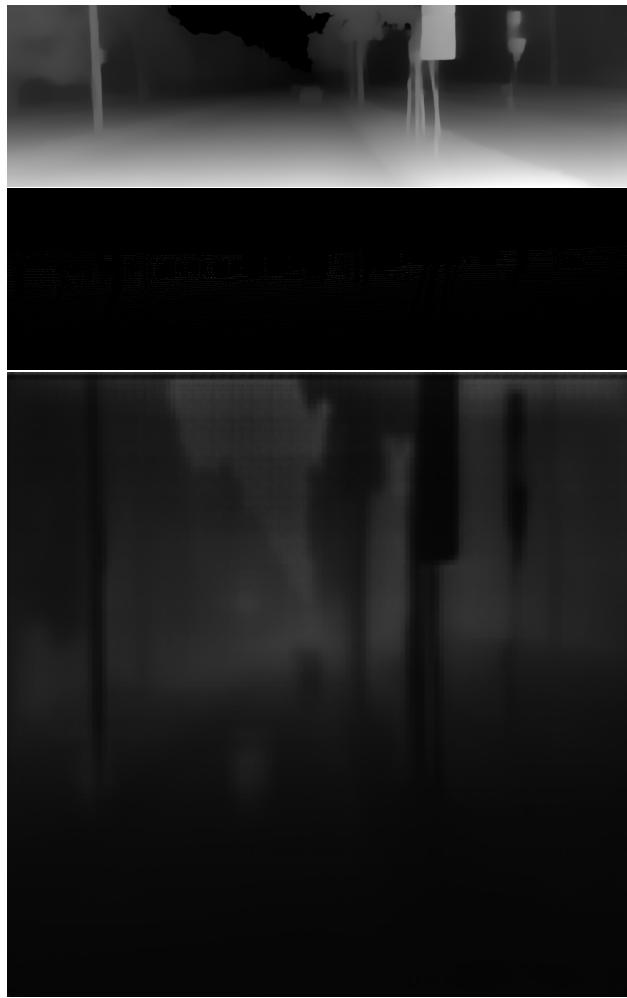


Figure 3.3: RGB LiDAR and predicted LiDAR.

3.3.3 Depth Anything v2 to LIDAR

Depth Anything v2, an enhancement over v1, demonstrated superior performance in predicting LIDAR intensity. This version incorporated improvements that resulted in more accurate and reliable depth predictions, leading to better LIDAR intensity estimates.



Figure 3.4: Depthanything v2 LiDAR and predicted LiDAR.

3.3.4 Metric Depth to LIDAR

In this approach, precise metric depth measurements were used to predict LIDAR intensity. This method significantly outperformed predictions based solely on RGB or depth maps, showcasing the effectiveness of metric depth information in enhancing LIDAR intensity predictions.

3.3.5 RGB plus BP Net to LIDAR

In this setup, RGB images were combined with depth maps generated by the BP Net model to predict LIDAR intensity. The integration of depth information from BP Net significantly improved the prediction accuracy compared to using RGB images alone. This method demonstrated enhanced performance across all evaluation metrics.

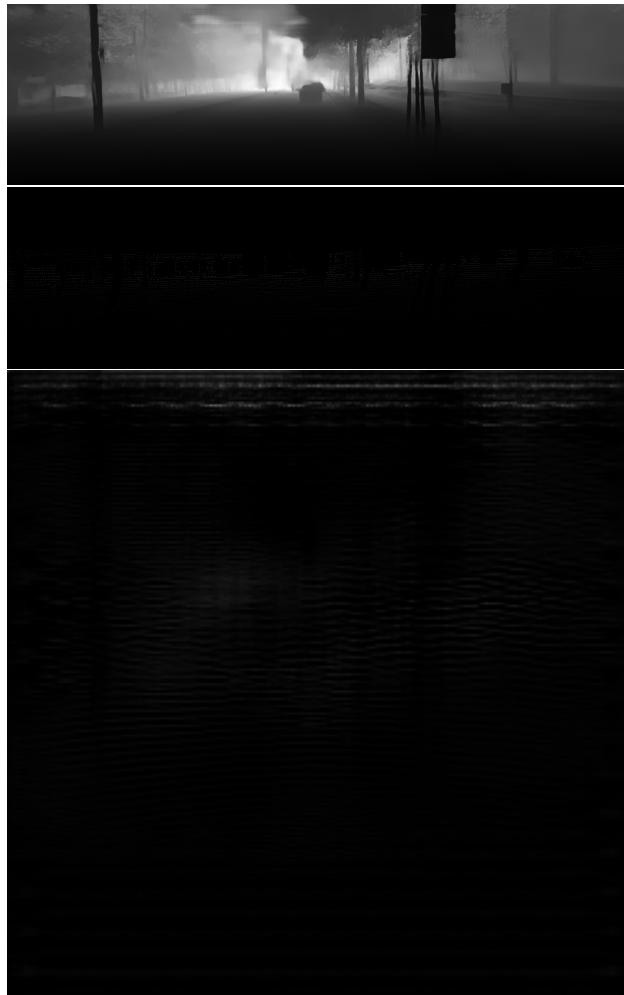


Figure 3.5: Metricdepth LiDAR and predicted LiDAR.

3.3.6 RGB plus Depth Anything v1 to LIDAR

Combining RGB images with depth maps from Depth Anything v1, the model achieved better performance than using RGB or depth alone. The fusion of RGB and depth data provided a more comprehensive input, resulting in improved LIDAR intensity predictions.

3.3.7 RGB plus Depth Anything v2 to LIDAR

With RGB images paired with depth maps from Depth Anything v2, the model achieved the highest performance among all setups. Depth Anything v2's enhanced depth predictions, combined with RGB data, resulted in highly accurate LIDAR intensity predictions.

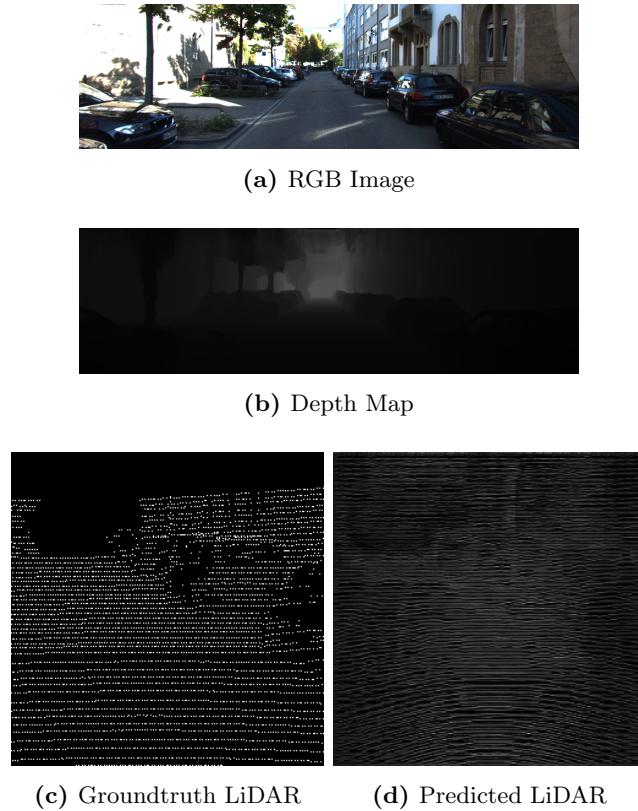


Figure 3.6: RGB, Depth, Predicted LiDAR, and Predicted LiDAR images for the BP Net setup.

3.3.8 RGB plus Metric Depth to LIDAR

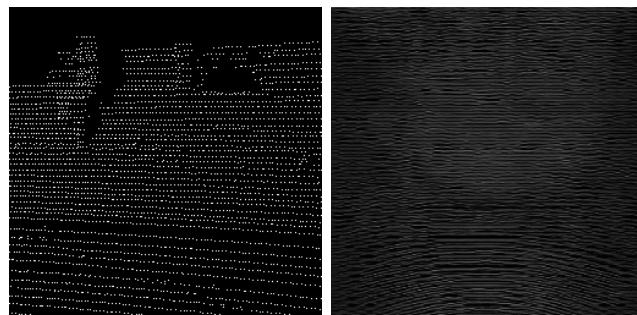
The combination of RGB images with precise metric depth measurements led to the most accurate predictions of LIDAR intensity. This setup utilized the strengths of both RGB data and metric depth, achieving the best results across all runs.



(a) RGB Image



(b) Depth Map



(c) Groundtruth LiDAR

(d) Predicted LiDAR

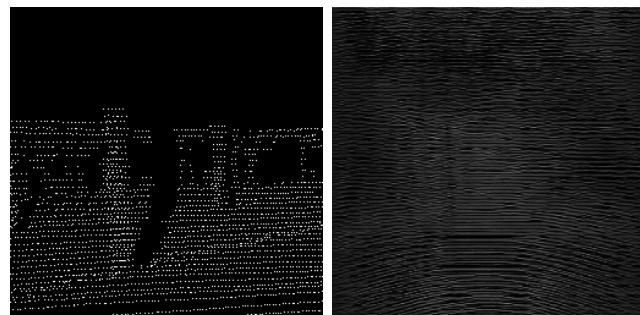
Figure 3.7: RGB, Depth, Predicted LiDAR, and Predicted LiDAR images for the Depth Anything v1 setup.



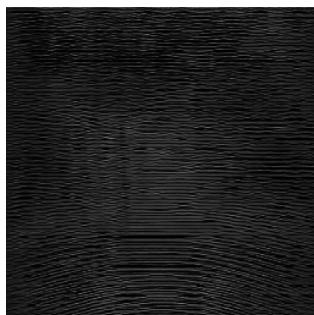
(a) RGB Image



(b) Depth Map



(c) Groundtruth LiDAR



(d) Predicted LiDAR

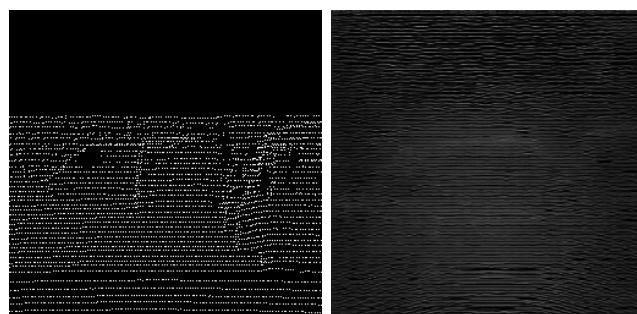
Figure 3.8: RGB, Depth, Predicted LiDAR, and Predicted LiDAR images for the Depth Anything v2 setup.



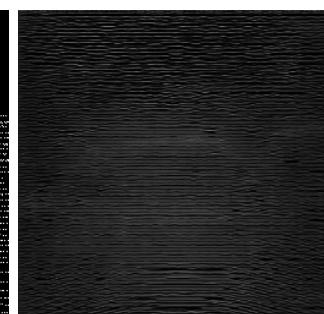
(a) RGB



(b) Depth



(c) Groundtruth LiDAR



(d) Predicted LiDAR

Figure 3.9: RGB Depthanything metric LiDAR and predicted LiDAR.

Chapter 4

Conclusion

The approaches with the rgb and depth only result in bad prediction. the bp

4.1 Appendix

4.1.1 RGB to LIDAR

The model trained using only RGB images to predict LiDAR intensity was evaluated. RGB alone provided sparse level of prediction, it lacked the accuracy required for high-precision applications. Here is a citation for the Depth Anything paper [3], its version 2 [2], and for the paper by Saxena et al. [1]

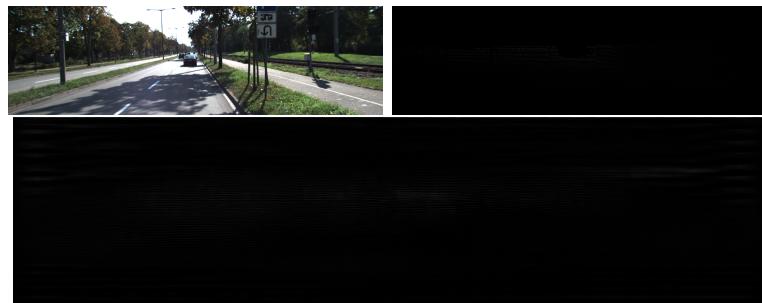


Figure 4.1: RGB LiDAR and predicted LiDAR.

4.2 References

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

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](https://doi.org/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.

List of Figures

3.1	RGB and created depth (BP, Deptanything v1,2 and metric v2)	6
3.2	RGB LiDAR and predicted LiDAR.	7
3.3	RGB LiDAR and predicted LiDAR.	8
3.4	Depthanything v2 LiDAR and predicted LiDAR.	9
3.5	Metricdepth LiDAR and predicted LiDAR.	10
3.6	RGB, Depth, Predicted LiDAR, and Predicted LiDAR images for the BP Net setup.	11
3.7	RGB, Depth, Predicted LiDAR, and Predicted LiDAR images for the Depth Anything v1 setup.	12
3.8	RGB, Depth, Predicted LiDAR, and Predicted LiDAR images for the Depth Anything v2 setup.	13
3.9	RGB Depthanything metric LiDAR and predicted LiDAR.	14
4.1	RGB LiDAR and predicted LiDAR.	15