

Predicting LIDAR Intensity from RGB and Depth Images

Project report in computer science for sensor simulation challenge

vorgelegt von

Carsten Schmotz

geb. am 23.09.1996 in Dachau

**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	3
1.1	Motivation	3
1.2	Contribution	3
1.3	Related Work	3
2	Predicting LiDAR Intensity	4
2.1	Setup	4
2.1.1	Bilateral Propagation Network (BP Net)	4
2.1.2	Depth Anything Models	4
2.1.3	Pix2pix Network	4
2.2	Implementation	5
2.2.1	Depth Map Generation	5
2.2.2	Model Training	6
2.3	Results	7
2.3.1	Depth BP Net to LiDAR	7
2.3.2	Depth Anything v1 to LiDAR	8
2.3.3	Depth Anything v2 to LiDAR	9
2.3.4	Metric Depth to LiDAR	10
2.3.5	RGB plus BP Net to LiDAR	11
2.3.6	RGB plus Depth Anything v1 to LiDAR	12
2.3.7	RGB plus Depth Anything v2 to LiDAR	13
2.3.8	RGB plus Metric Depth to LiDAR	13
3	Conclusion	16
3.1	Appendix	16
3.1.1	RGB to LiDAR	16
3.2	References	16

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

Predicting LiDAR Intensity

2.1 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.

2.1.1 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.

2.1.2 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.

2.1.3 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 2.3.

2.2 Implementation

2.2.1 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 2.1) 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 2.1: RGB and created depth (BP, Deptanything v1,2 and metric v2)

2.2.2 Model Training

- **pix2pix Configuration:** The pix2pix architecture was adapted to accept RGB images and depth maps (4 dim.) as inputs to be able to scan input channel separately and to avoid loss by overlaying rgb and depth images. The unmodified pix2pix was trained to predict LiDAR intensity values based on depthmaps inputs. The 4 dimensional pix2pix was given the input of rgb plus depthmaps.
- **Training Process:** The training parameters such as learning rate 0.0002000, batch size 1, and number of epochs 20 were set. The model was trained using the prepared dataset in the configuration 1000 pictures divided in 800 train, 100 val, 100 test folders. The input is different for the sole depth input and the rgbd input due to not equal running scripts for dividing. This was discovered after the training and testing runs. Both versions were executed with masking out the pixel value -1 to reduce the areas which are not of interest. The input

was crop to a size of 256 time 256 but not scale.

2.3 Results

This section presents the results of the experiments using different setups to predict LiDAR intensity. It was evaluated the performance of several methods, including the seperate depths from BP Net, Depth Anything v1 and v2, as well as metric depth estimation to LiDAR. The other four are RGB plus each of these depths to LiDAR. The returning pictures are crop and the full verions a used as comparison. The crop version is the right part of the full resolution (1216×352) images. For the 4 dimensional runs the cropped LiDAR groundtruth for better visualisation because this pix2pix version gives these extra outputs. The results from eight runs are summarized below:

2.3.1 Depth BP Net to LiDAR

Using only the depth maps generated by BP Net (Figure 2.2), without accompanying RGB images, the model was trained to predict LiDAR intensity. This approach yielded shows some visible prediction.

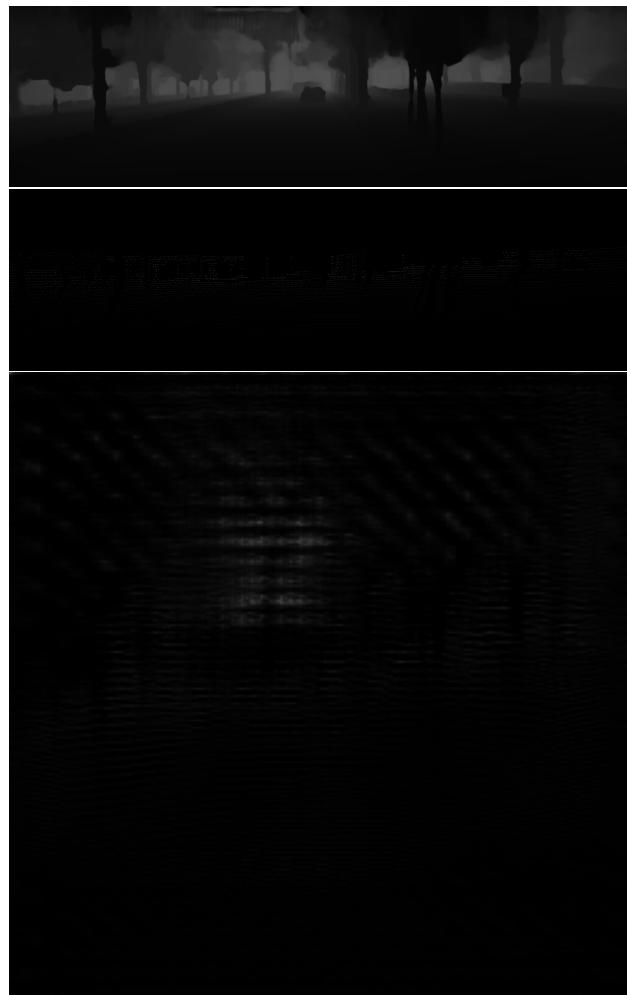


Figure 2.2: RGB LiDAR and predicted LiDAR.

2.3.2 Depth Anything v1 to LiDAR

The Depth Anything v1 model (Figure 2.3) provided initial depth estimations using large-scale unlabeled data. When used to predict LiDAR intensity, this model showed results that are more familiar with depth maps and not LiDAR intensity.

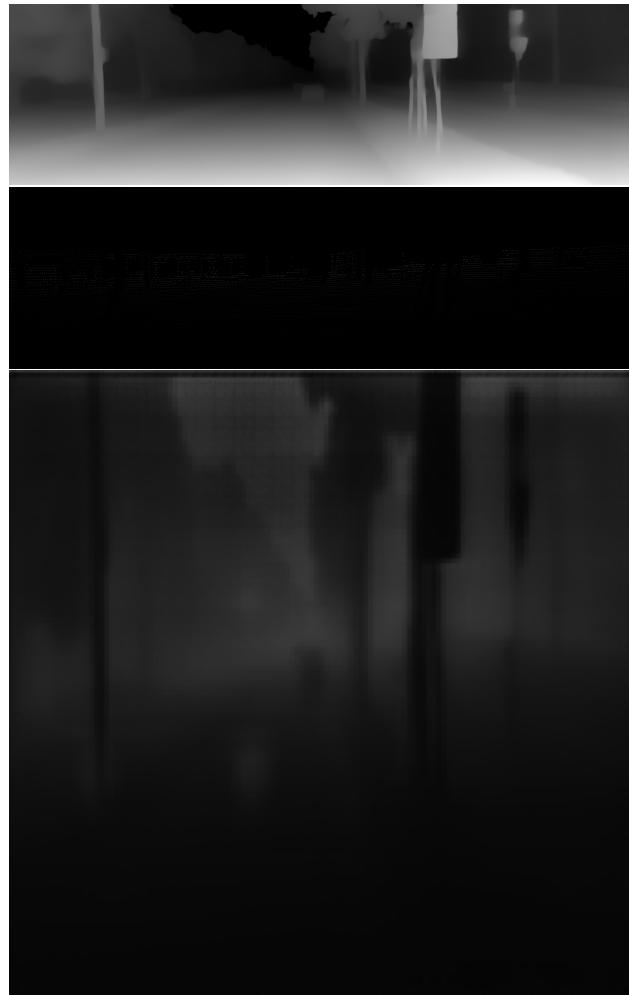


Figure 2.3: RGB LiDAR and predicted LiDAR.

2.3.3 Depth Anything v2 to LiDAR

The Depth Anything v2 (Figure 2.4) shows almost the same result like v1. It could be possible that pix2pix tried to learn LiDAR to depth but the datasets and the training settings were correct for that.



Figure 2.4: Depthanything v2 LiDAR and predicted LiDAR.

2.3.4 Metric Depth to LiDAR

In this approach, precise metric depth measurements (Figure 2.5) were used to predict LiDAR intensity. This method significantly shows some good prediction of the LiDAR intensity.

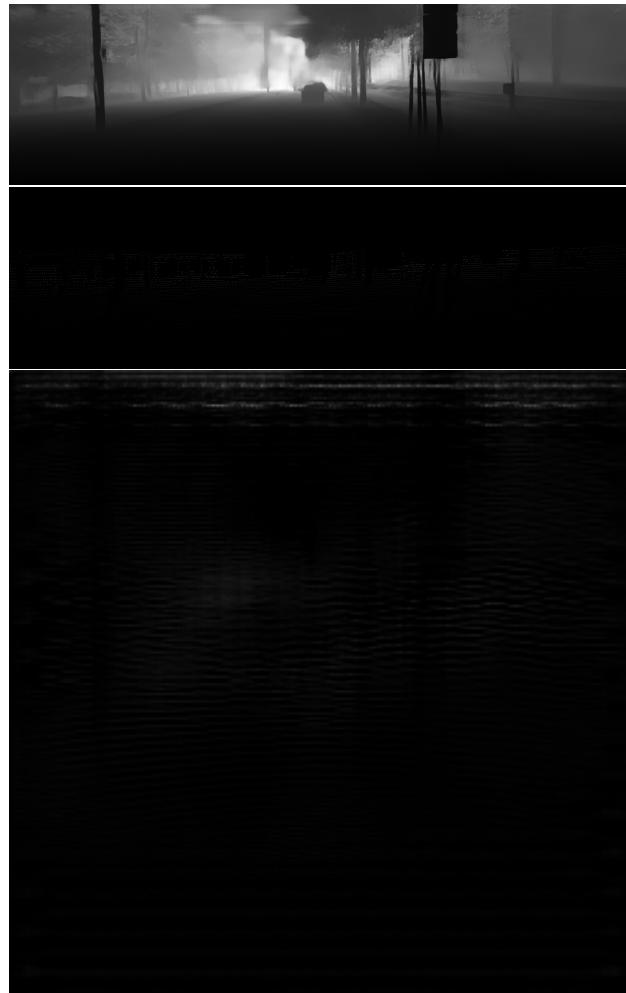


Figure 2.5: Metricdepth LiDAR and predicted LiDAR.

2.3.5 RGB plus BP Net to LiDAR

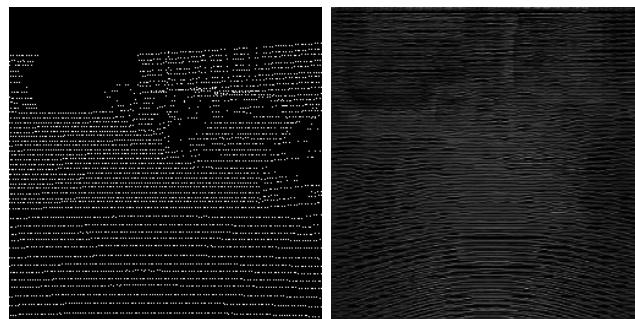
In this setup (Figure 2.6), RGB images were combined with depth maps generated by the BP Net model to predict LiDAR intensity. The integration of depth information from BP Net significantly displays some correct prediction accuracy compared to using Depth maps images alone. The cars for instance are barely visible and the background is almost filled wrong perhaps due to the masking.



(a) RGB Image



(b) Depth Map



(c) Groundtruth LiDAR (d) Predicted LiDAR

Figure 2.6: RGB, Depth, Predicted LiDAR, and Predicted LiDAR images for the BP Net setup.

2.3.6 RGB plus Depth Anything v1 to LiDAR

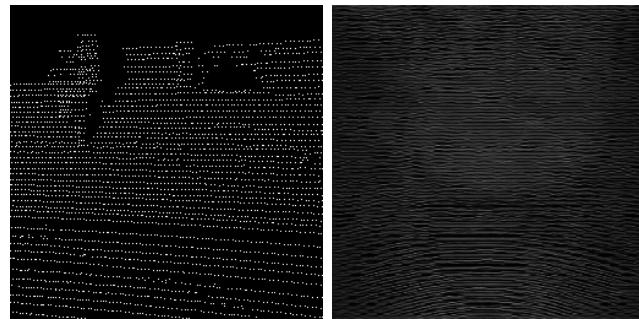
Combining RGB images with depth maps from Depth Anything v1 (Figure 2.7), the model achieved better performance than using depth alone. The fusion of RGB and depth data provided a more comprehensive input, resulting in an LiDAR intensity but it show reselembles to the groundtruth. The end of the street is cleary visible.



(a) RGB Image



(b) Depth Map



(c) Groundtruth LiDAR

(d) Predicted LiDAR

Figure 2.7: RGB, Depth, Predicted LiDAR, and Predicted LiDAR images for the Depth Anything v1 setup.

2.3.7 RGB plus Depth Anything v2 to LiDAR

RGB images paired with depth maps from Depth Anything v2 (Figure 2.4) show a better prediction than v1. The sign in the groundtruth is more visible due to better depthmaps resolution of the depthanything v2.

2.3.8 RGB plus Metric Depth to LiDAR

The combination of RGB images with precise metric (Figure 2.9) 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. The contours of the cars on the right side are visible.



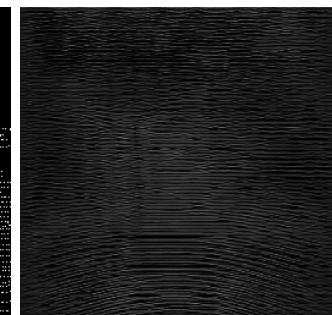
(a) RGB Image



(b) Depth Map



(c) Groundtruth LiDAR



(d) Predicted LiDAR

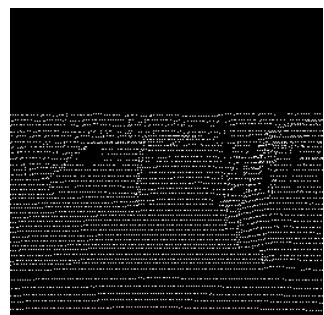
Figure 2.8: RGB, Depth, Predicted LiDAR, and Predicted LiDAR images for the Depth Anything v2 setup.



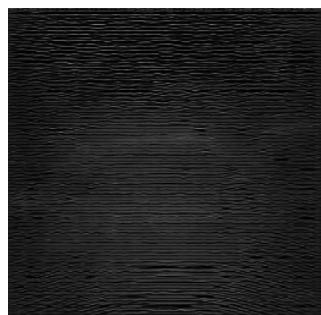
(a) RGB



(b) Depth



(c) Groundtruth LiDAR



(d) Predicted LiDAR

Figure 2.9: RGB Depthanything metric LiDAR and predicted LiDAR.

Chapter 3

Conclusion

In conclusion, the results of different depthmaps against LiDAR are showing bad prediction for the for most of the sole depthmaps input. The combine inputs are performing better on the prediction. The comparions are not fully acceptalbe because there were problems with the different output images and mostly not machting samples. Furhtermore did the provide script not work for this installment because the cropped output picture should be in pairs of 4 be stited togehter to get the full image. This leads to an difficult evaltion on the data.

Overall there are some differnece to be notice dephťanything v2 performces better dann v1 and the metric depth is better than both. The BP Net shows also some good result. A better approch is to use the same output images and to run without masking for a more reliable result. Furthermore is the missing of loss function due to not working scripts a big disadvantage for evalution.

3.1 Appendix

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

3.2 References

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

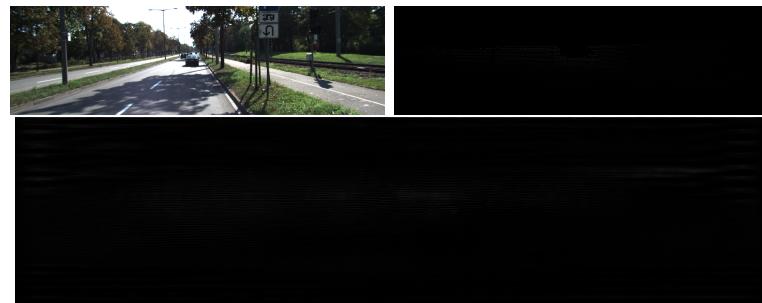


Figure 3.1: RGB LiDAR and predicted LiDAR.

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.