PriorDepth

Advanced Topics in 3D Computer Vision Praktikum

Team: Rahul, Halil, Konstantin
Advisor: Patrick Ruhkamp

Final Presentation

PriorDepth - Problem Definition

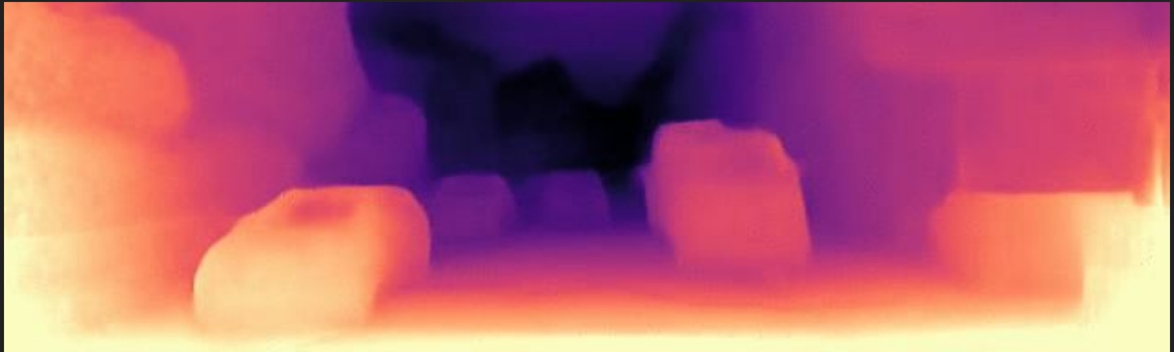
Task 3: Unsupervised Depth Estimation from Sparse Spatial-Temporal Priors

- Building Depth Estimation and Ego-Motion Estimation Pipeline, applicable for Outdoor and Indoor Cameras
- Accurate Depth Estimation for slow and fast moving Cameras, including Hand Held Cameras
- For the base networks we use **MonoDepth2**[1] and **KeypointNet2D**[2]
- The pipeline structure is adapted from KeypointNet3D[3]
- Pose Estimation for losses should be computed with a **Differentiable Ego-Motion Estimator.** for end to end training.

1.Godard, C. et al. (2019) Digging Into Self-Supervised Monocular Depth Estimation (ICCV)
2.Tang, J. et al. (2019) Neural Outlier Rejection for Self-Supervised Keypoint Learning (ICLR)
3.Jiexiong Tang; Self-Supervised 3D Keypoint Learning for Ego-motion Estimation, 2020

PriorDepth - Motivation Robust Depth Estimation

Depth Estimations are reproduced from MonoDepth2[1] on Kitti[2]



1.Godard, C. et al. (2019) Digging Into Self-Supervised Monocular Depth Estimation (ICCV)
2.Geiger, A., et al. “Vision Meets Robotics: The KITTI Dataset.” The International Journal of Robotics Research, vol. 32, no. 11, Sept. 2013

PriorDepth - Motivation for Robust Depth Estimation

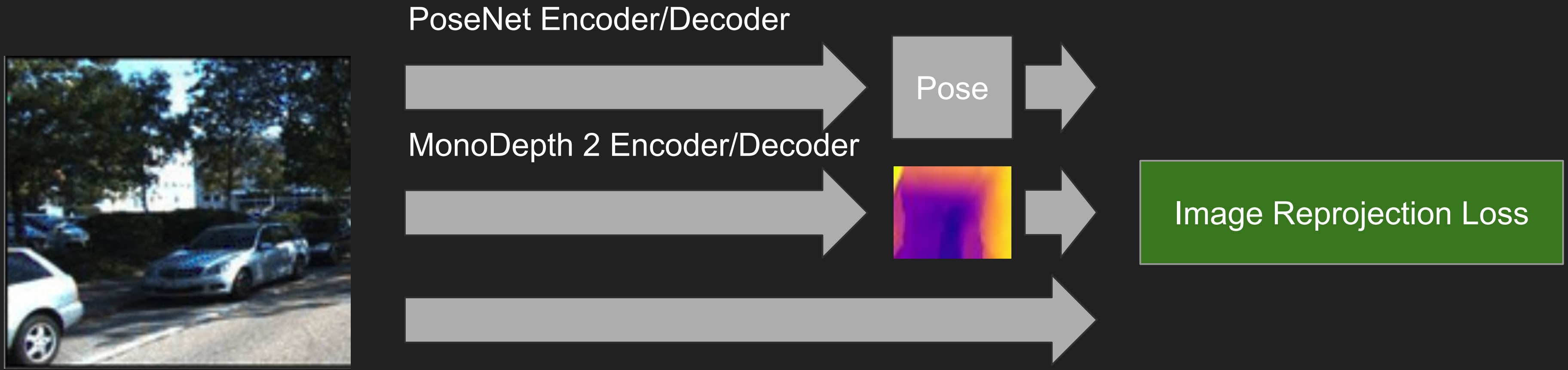
Depth Estimation Circumstances

Outdoor		Indoor
Steady Cameras	VS.	Hand Held Camera
Similar Pace		Diverse Pace

1.Godard, C. et al. (2019) Digging Into Self-Supervised Monocular Depth Estimation (ICCV)

Priordepth - MonoDepth2

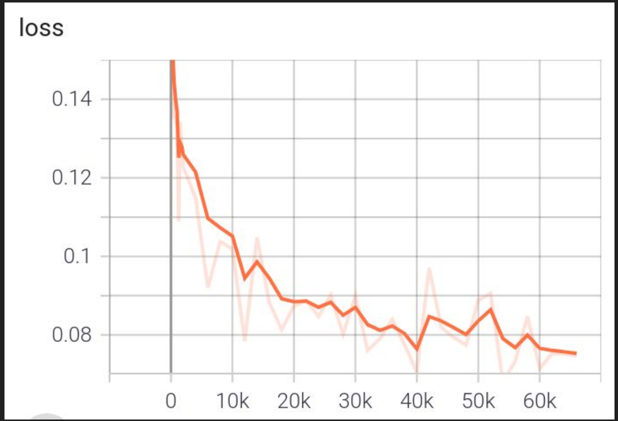
__Monodepth Structure__



1. Godard, C. et al. (2019) Digging Into Self-Supervised Monocular Depth Estimation (ICCV)
2. Kendall, A. et al. (PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization

Priordepth - MonoDepth2

Monodepth Reproduction



Training Settings:

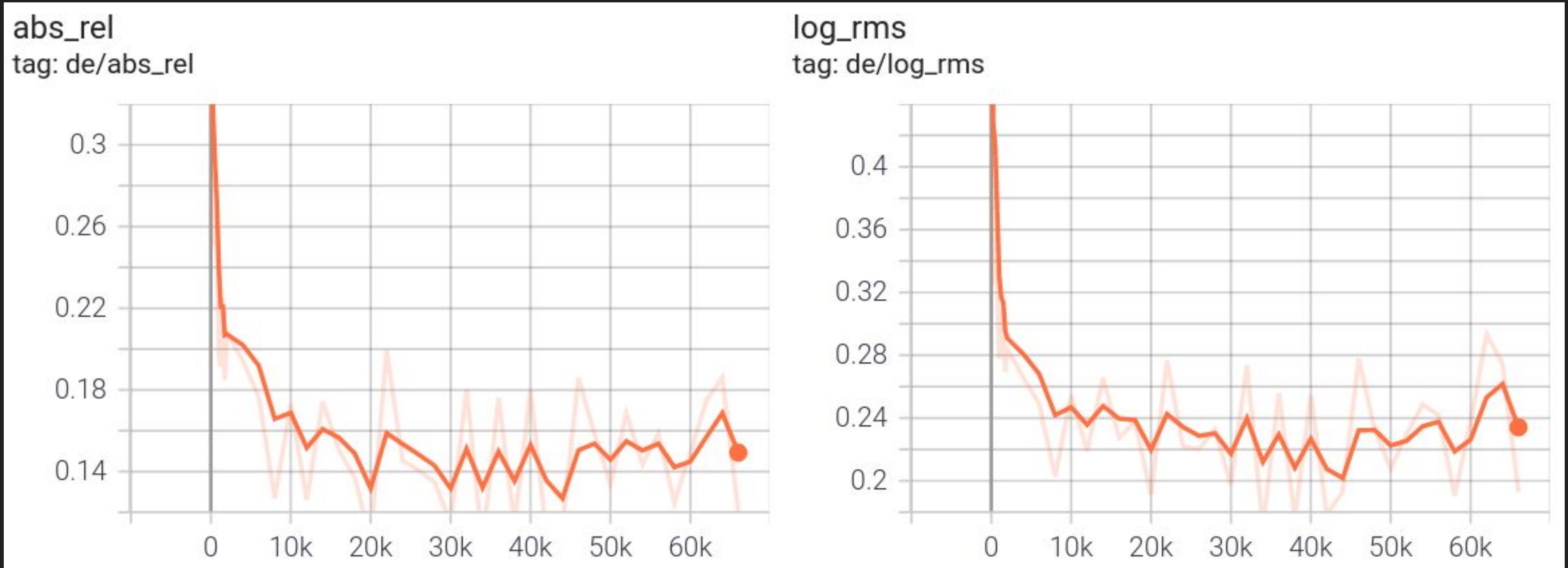
- 20 Epochs, adam optimizer, Batch size = 12, Lr = 1e-3
- Eigen Zhou split with around 44 000 images 10% validation 90% training
- ImageNet Pretrained (default)

	Abs Rel	Sq Rel	RMSE	RMSE log
Reproduction Model	0.1493	0.87	5.016	0.1927
Monodepth2	0.132	1.044	4.872	0.210

1. Godard, C. et al. (2019) Digging Into Self-Supervised Monocular Depth Estimation (ICCV)

Priordepth - Monodepth 2

Monodepth Reproduction - Fast Convergence



1.Godard, C. et al. (2019) Digging Into Self-Supervised Monocular Depth Estimation (ICCV)

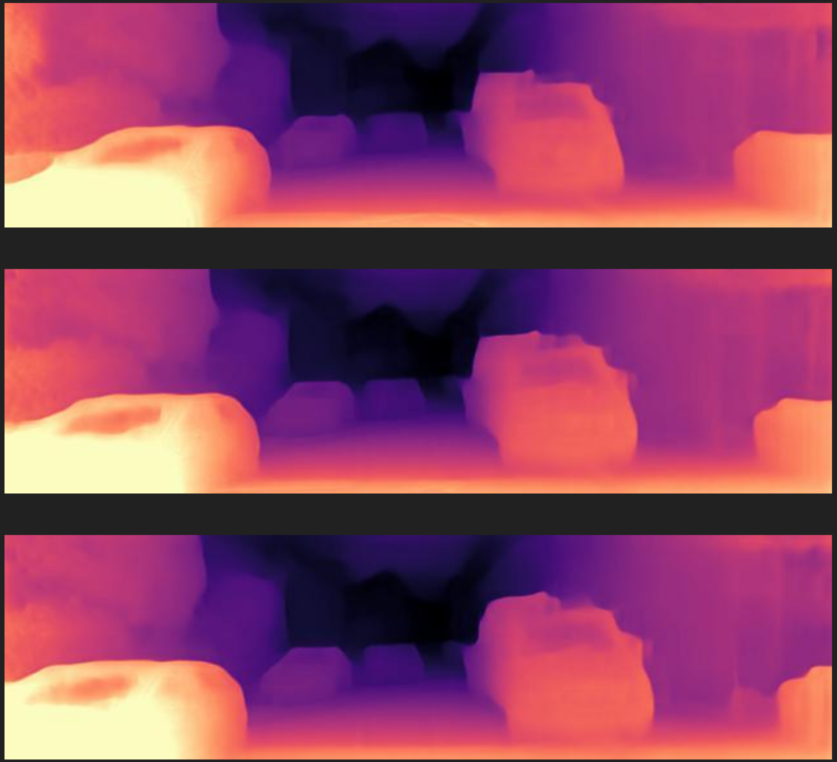
Priordepth - Monodepth 2

Monodepth Drawbacks

- Monodepth2[1] uses PoseNet

- Trained on kitti - slow driving in input frames

- PoseNet Overfitting -> Pose is used in Loss Calculation
- Unuseable for in-door data and most other datasets



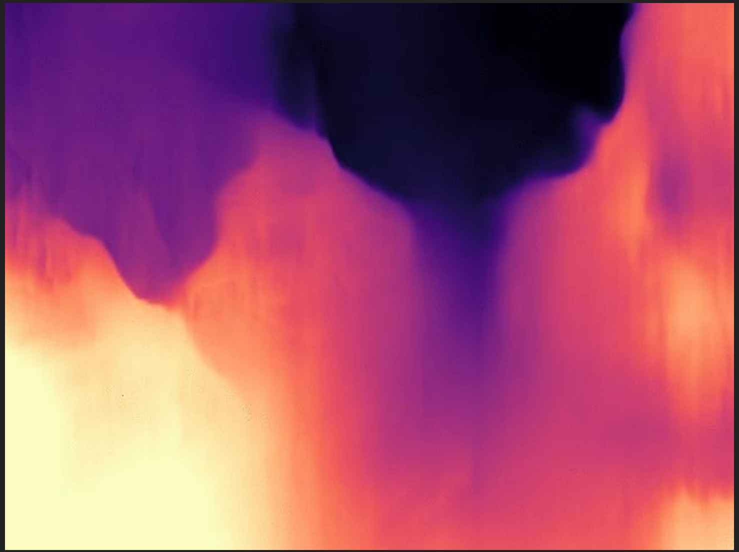
3 consecutive Depth estimations with own reproduced Monodepth 2

1.Godard, C. et al. (2019) Digging Into Self-Supervised Monocular Depth Estimation (ICCV)

Priordepth - Monodepth 2

Monodepth Drawbacks

RGBD Indoor Dataset[2]

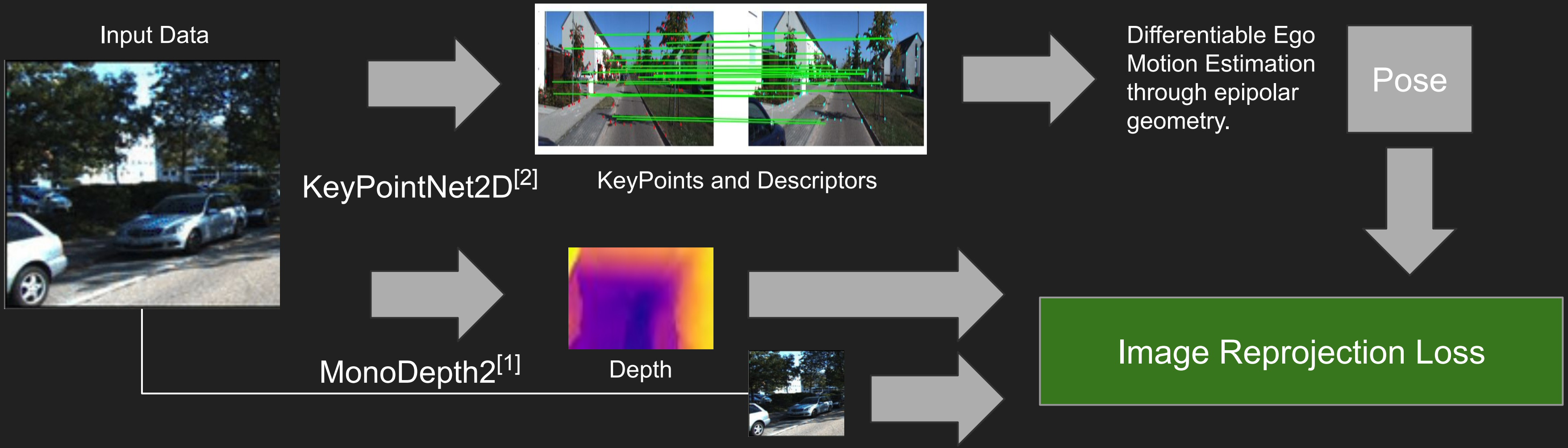


Depth Estimation[1]

1. Godard, C. et al. (2019) Digging Into Self-Supervised Monocular Depth Estimation (ICCV)
2. J. Sturm and N. et al. (2012) A Benchmark for the Evaluation of RGB-D SLAM Systems

PriorDepth – General Idea

- Unsupervised Depth and Ego-Motion Estimation with Temporal Consecutive Monocular View using Keypoints
- Goal: Predict better Poses to formulate better reprojection loss, improving depth estimation for indoor scenes and close ranges

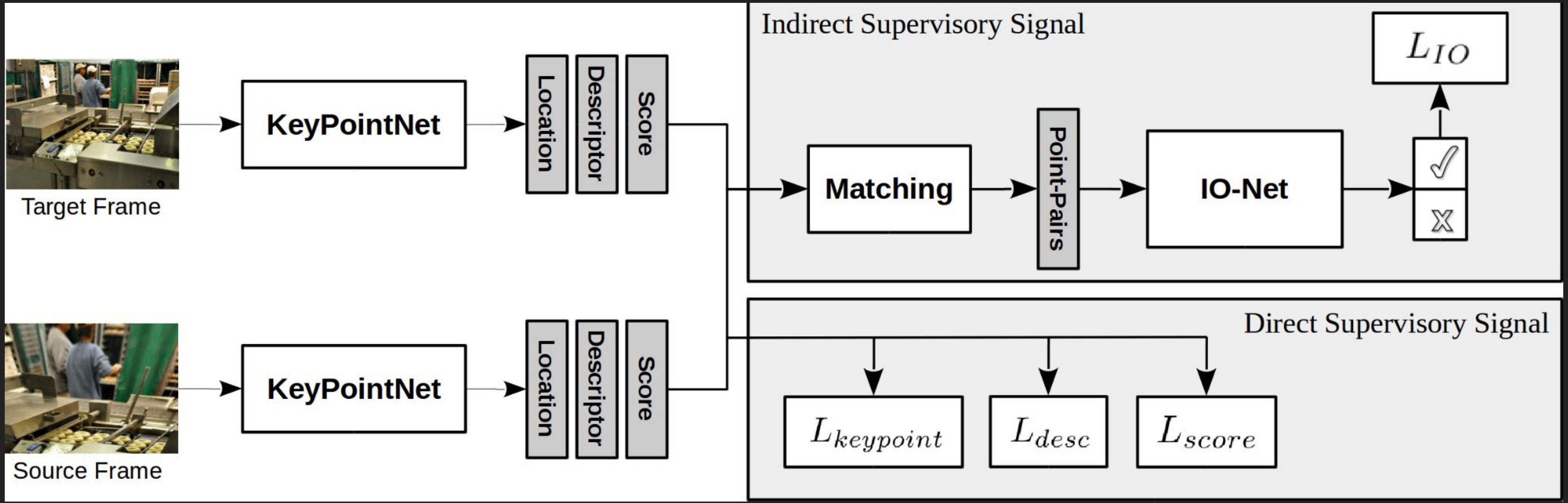


1.Godard, C. et al. (2019) Digging Into Self-Supervised Monocular Depth Estimation (ICCV)
2.Tang, J. et al. (2019) Neural Outlier Rejection for Self-Supervised Keypoint Learning (ICLR)

PriorDepth – KP2D[1]

Extracts the keypoints, descriptors and the scores

Improve Outlier Rejection



Keypoint Loss
Distance between the target keypoint and warped source keypoint

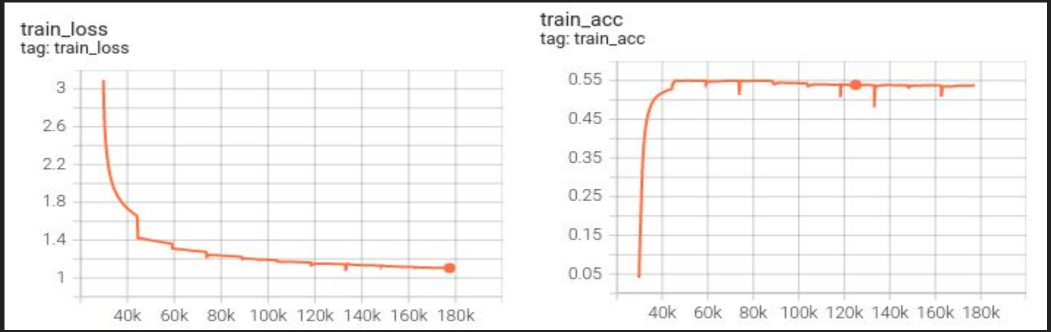
Descriptor Loss
Per pixel Triplet Loss on distance between the descriptors
+ve and -ve samples from keypoint correspondences between the images

Score Loss
Minimize the distance between scores of keypoint pairs + min./max. the average scores of keypoint pair

[1] Tang, Jiexiong, et al. "Neural Outlier Rejection for Self-Supervised Keypoint Learning." ArXiv:1912.10615

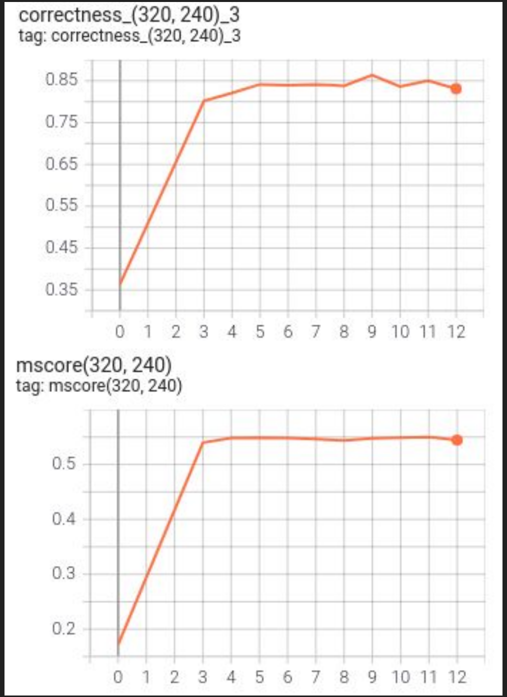
PriorDepth – KP2D

Training on COCO[1] 2017 (Train set)



Validation Metrics	Our Training (12 epochs)	Results from the paper (50 epochs)	Progress
C1	0.493	0.593	↑
C3	0.831	0.867	
C5	0.893	0.91	
Matching Score	0.544	0.546	
Repeatability	0.660	0.687	
Localization	0.913	0.892	↓

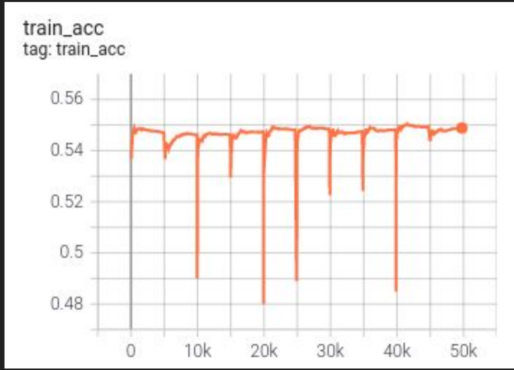
Validation on HPatches[2]



[1]Tsung-Yi Lin, et al. "Microsoft COCO: Common Objects in Context." (2015).
[2]Vassileios Balntas, et al. "HPatches: A benchmark and evaluation of handcrafted and learned local descriptors." (2017).

PriorDepth – KP2D

Pretraining on KITTI[1] - Eigen Zhou split



Help the network navigate the domain shift with pre-training

Use model that performs well on KITTI to plug into KP3D baseline model and freeze the network for initial configurations.

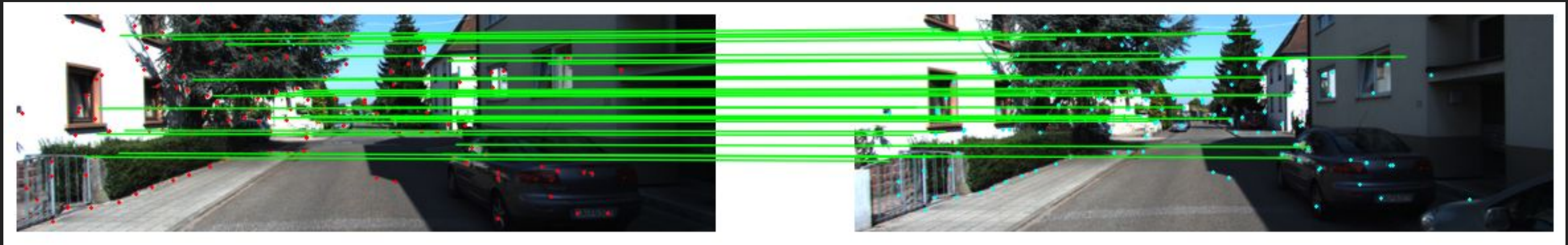
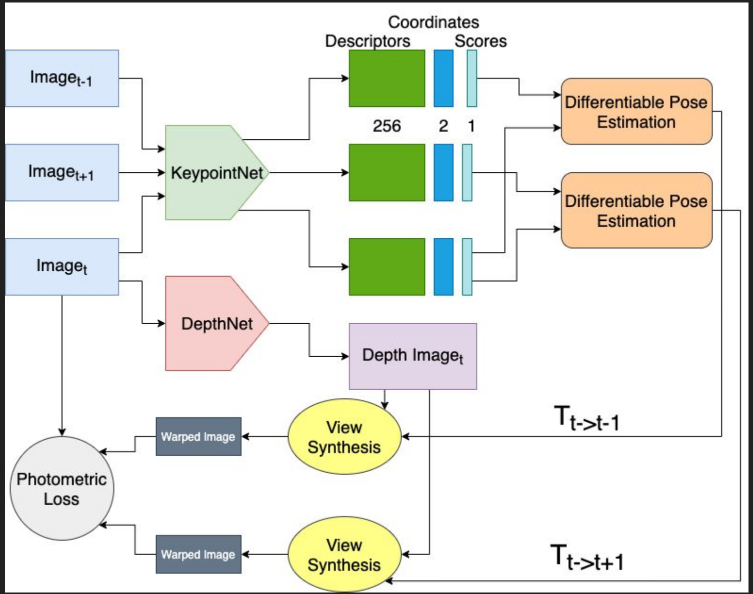


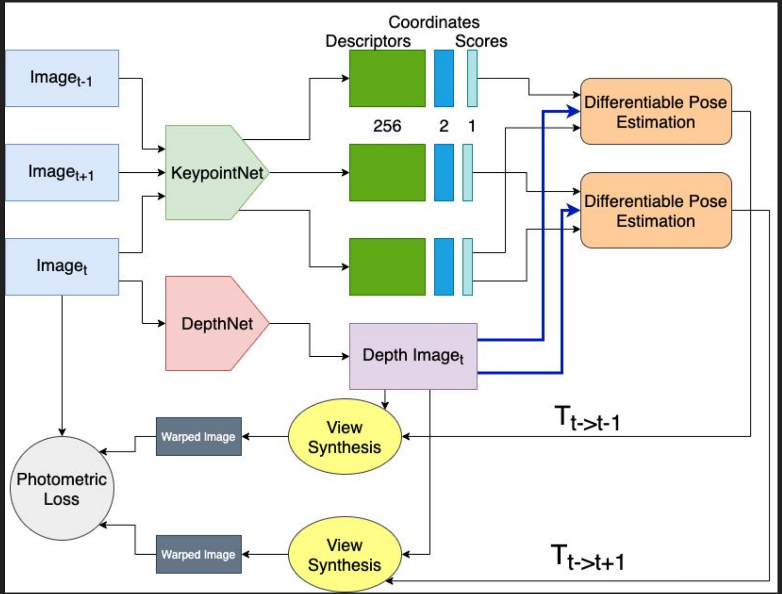
Figure - Visualisation of the matched keypoints on KITTI from KeypointNet

[1] Geiger, A., et al. “Vision Meets Robotics: The KITTI Dataset.” The International Journal of Robotics Research, vol. 32, no. 11, Sept. 2013

Priordepth - KP3D[1]



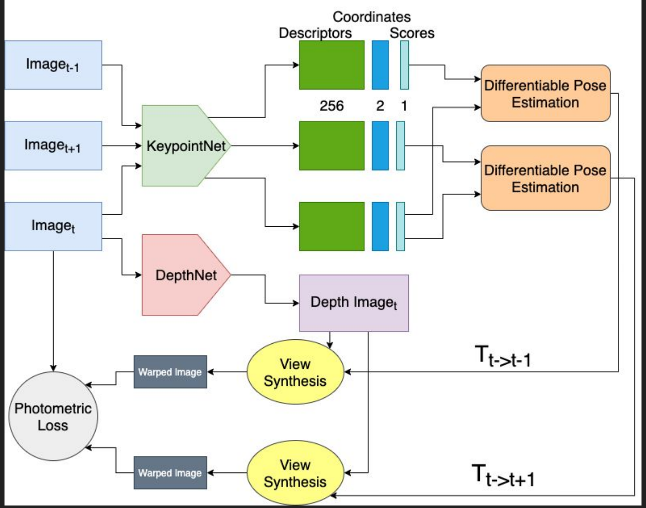
PriorDepth - Pose with Essential Matrix



PriorDepth - Pose with Perspective-n-Point

[1] Tang, Jiexiong, et al. “Self-Supervised 3D Keypoint Learning for Ego-Motion Estimation”

PriorDepth - KP3D



Possible Training Configurations

Train DepthNet

UnFrozen KeyPointNet

- Homographic warping
- Homographic warping + Perspective Warping

Frozen KeyPointNet

Pose Estimator

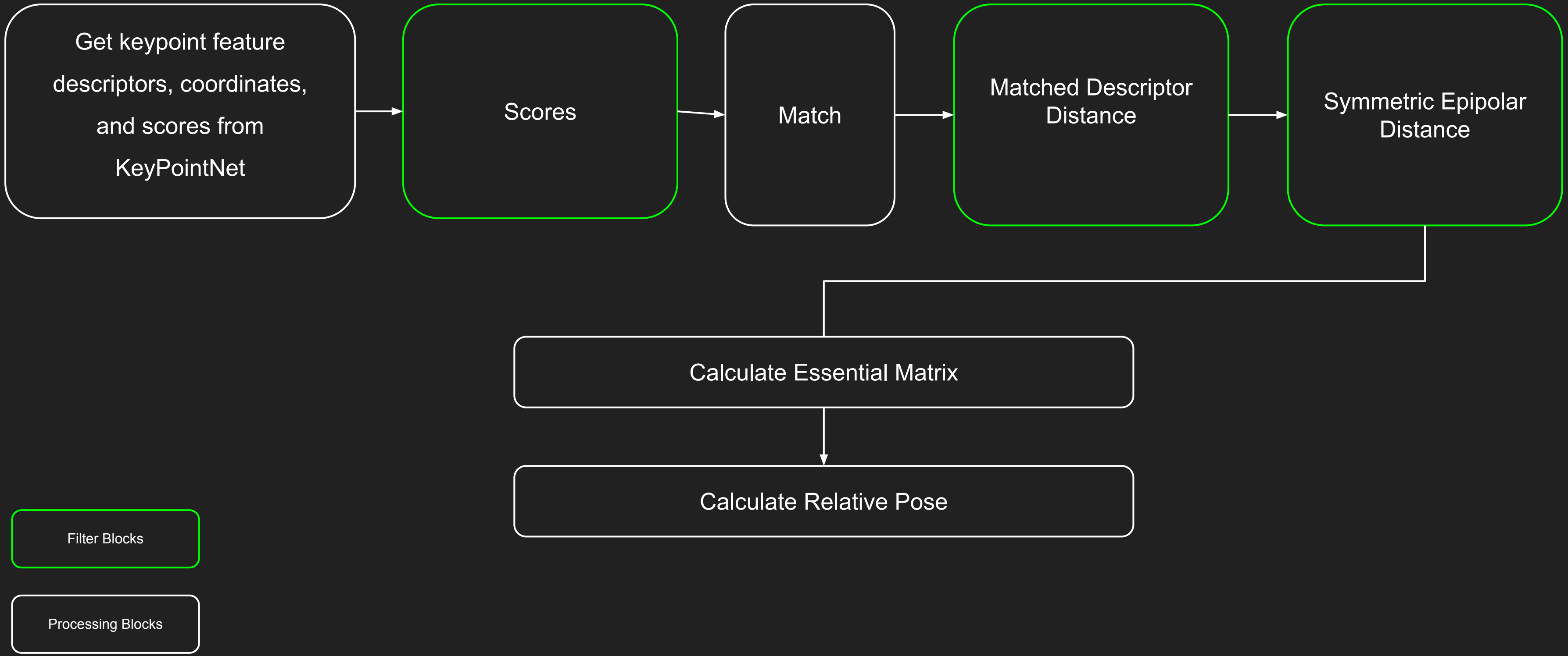
- Essential Matrix based Pose
- PnP based Pose

As Proposed in KP2D[1]

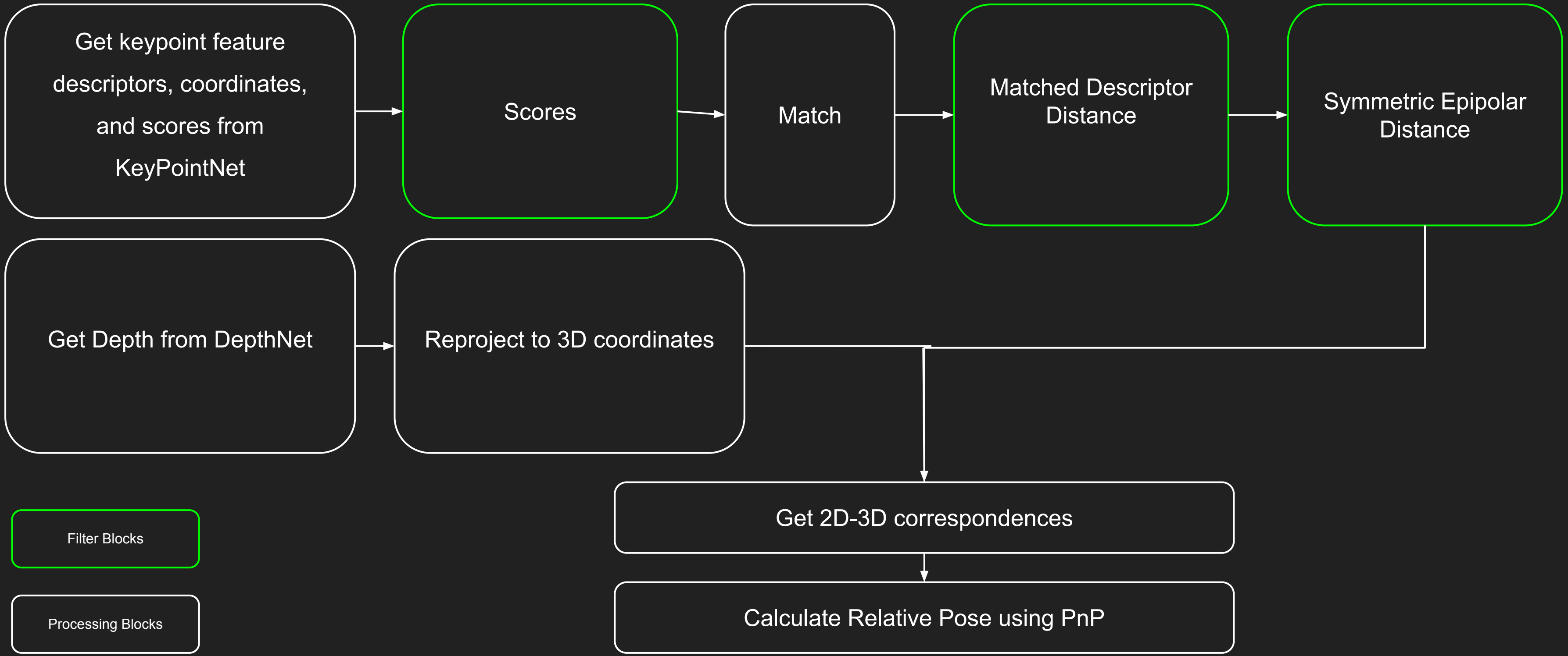
Our Extensions

[1] Tang, Jiexiong, et al. “Neural Outlier Rejection for Self-Supervised Keypoint Learning.” ArXiv:1912.10615

PriorDepth - Pose from Essential Matrix Expanded

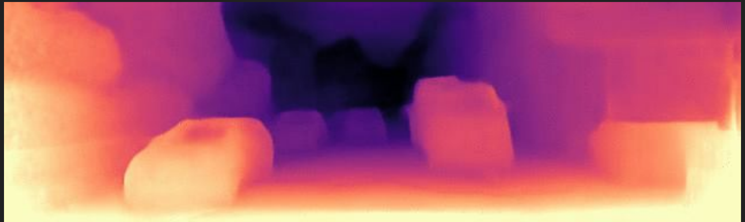


PriorDepth - Pose from PnP Expanded

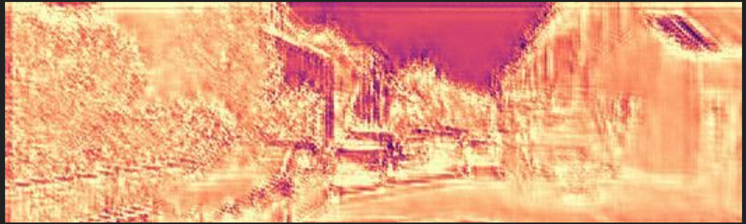


PriorDepth - Results and Challenges

PriorDepth - KITTI

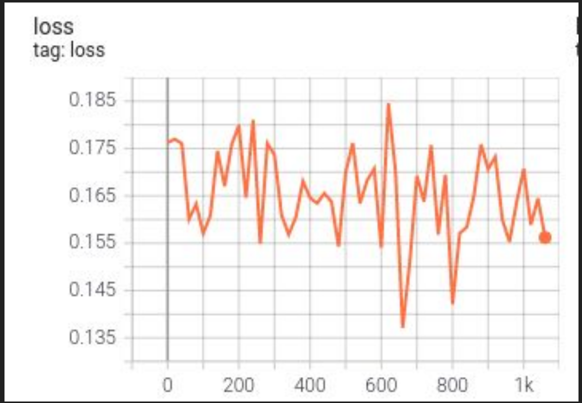


DepthNet[1] trained with PoseNet



Our Implementation based on Differential Pose Estimation

Inaccurate Depth is also reflected in the training curves.



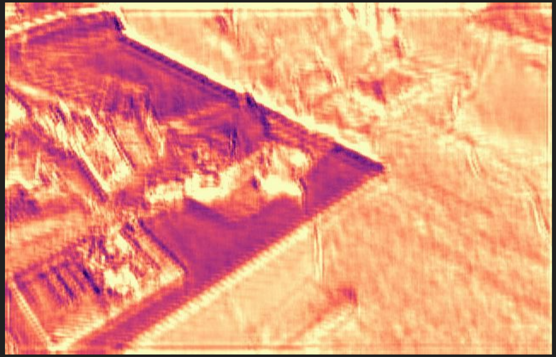
1.Godard, C. et al. (2019) Digging Into Self-Supervised Monocular Depth Estimation (ICCV)

PriorDepth - Results and Challenges

PriorDepth - Indoor Data

DepthNet trained with PoseNet

Original[1]

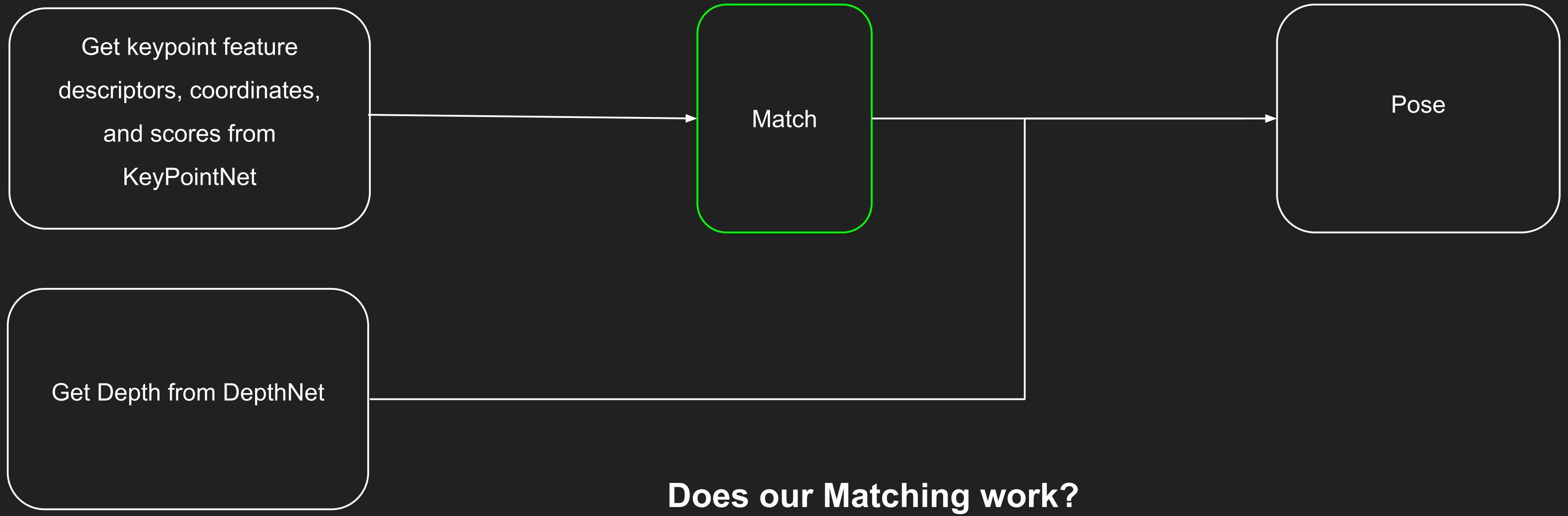


Our Implementation based on Differential Pose Estimation

1. J. Sturm and N. et al. (2012) A Benchmark for the Evaluation of RGB-D SLAM Systems

PriorDepth

Debug Process

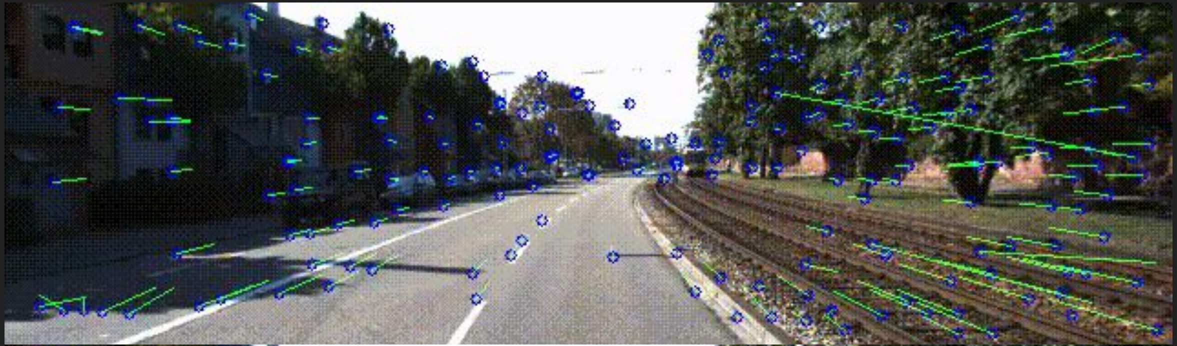


Does our Matching work?

PriorDepth

Debug Process

Before Filtering



After Filtering

Matching Works! Our keypoint filters also work.

Let’s check for warped image from calculated poses

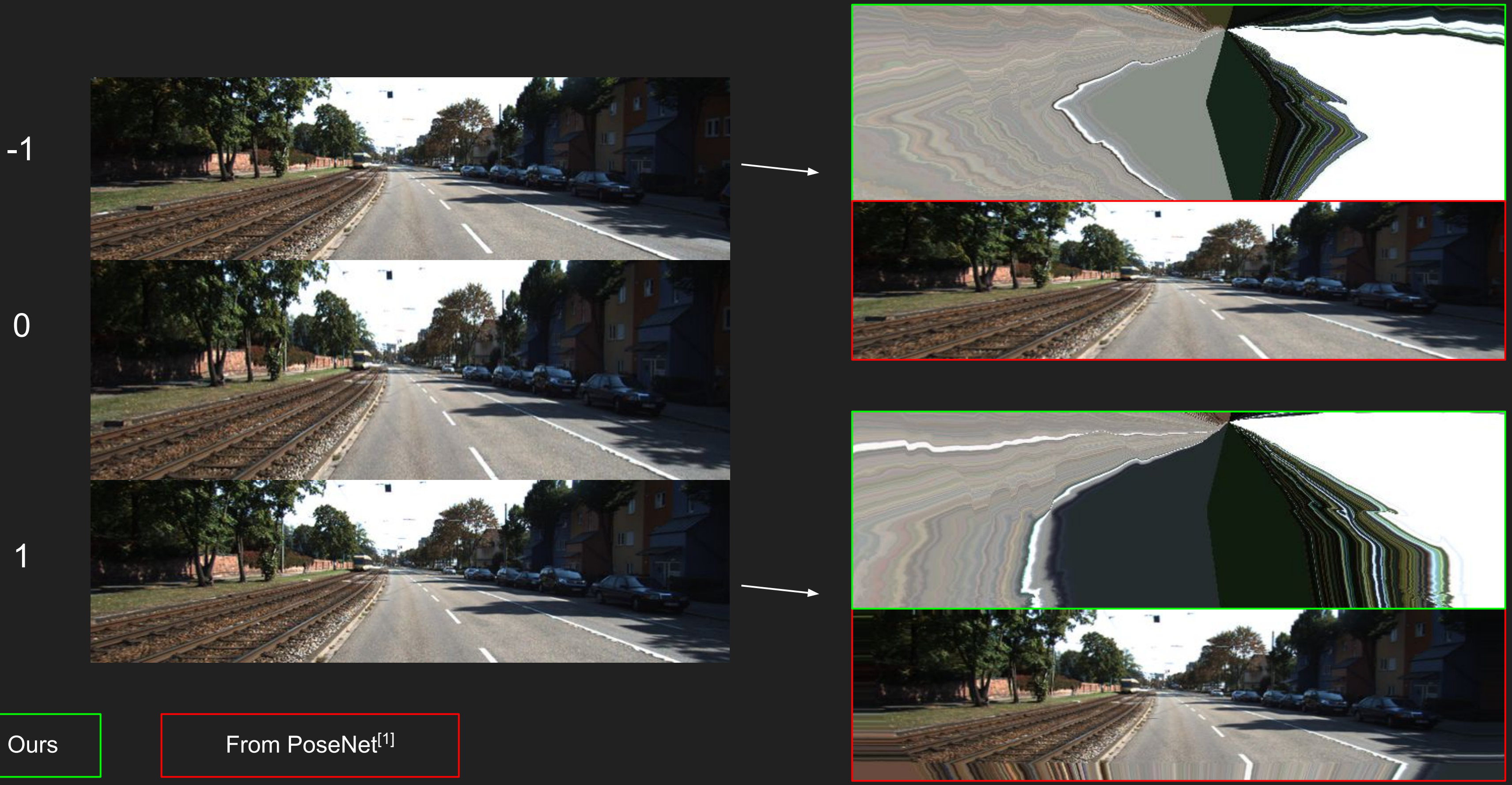
PriorDepth - 2D & 3D homography warping

Ours

From PoseNet[1]

1. Godard, C. et ai. (2019) Digging Into Self-Supervised Monocular Depth Estimation (ICCV)

PriorDepth - 2D & 3D homography warping PnP



Ours

From PoseNet[1]

1. Godard, C. et al. (2019) Digging Into Self-Supervised Monocular Depth Estimation (ICCV)

PriorDepth - Summary

- We tried to build an end to end differentiable pipeline for robust depth and ego motion estimation
- The loss values did not converge on training and the visualisations also showed the network could no train.
- On further debugging, we found that the matching and subsequent keypoint filters worked.
- After subsequent trials, it was found that the pose calculations were not accurate due to errors in estimating fundamental matrix.
- Due to inaccurate pose estimations, warping failed and hence, DepthNet was unable to train itself.

Thank you for the attention!

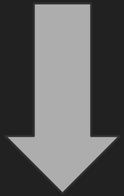
Questions?

Priordepth - Monodepth 2

Monodepth Drawbacks

PoseNet Overfitting trained on Kitti

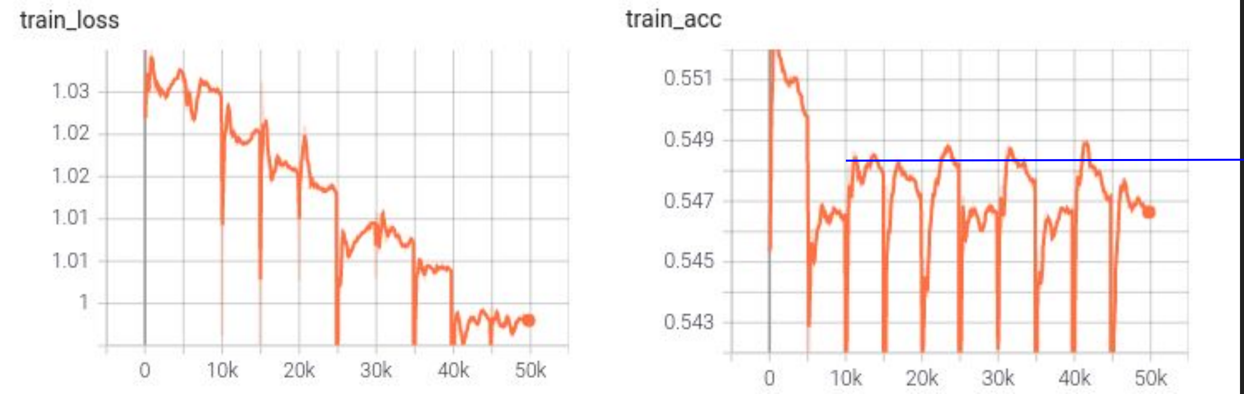
1. Car often slowly moving (before curves)
2. Often with constant pace and straight line



Monodepth produces bad outputs for fast camera movement like handheld cameras

PriorDepth – KP2D

Pretraining on KITTI[4] - Eigen Zhou split



Help the network navigate the domain shift with pre-training

Use model that performs well on KITTI to plug into KP3D baseline model and freeze the network.

Figure - Visualisation of the matched keypoints on KITTI from KeypointNet

[4] Geiger, A., et al. “Vision Meets Robotics: The KITTI Dataset.” The International Journal of Robotics Research, vol. 32, no. 11, Sept. 2013

Priordepth - KeypointNet 2D

KeypointNet 2D Drawbacks

PriorDepth - Pose from Essential Matrix

- Initial keypoint feature descriptors, coordinates, and scores from KP2D
- Filter out keypoints based on:
  - Score threshold
  - Descriptor-distance threshold
  - Distance on epipolar line
- Compute pose from keypoints left and essential matrix

1. Godard, C. et alii. (2019) Digging Into Self-Supervised Monocular Depth Estimation (ICCV)
2. J. Sturm and N. et alii. (2012) A Benchmark for the Evaluation of RGB-D SLAM Systems

PriorDepth - Pose from Perspective-n-Points

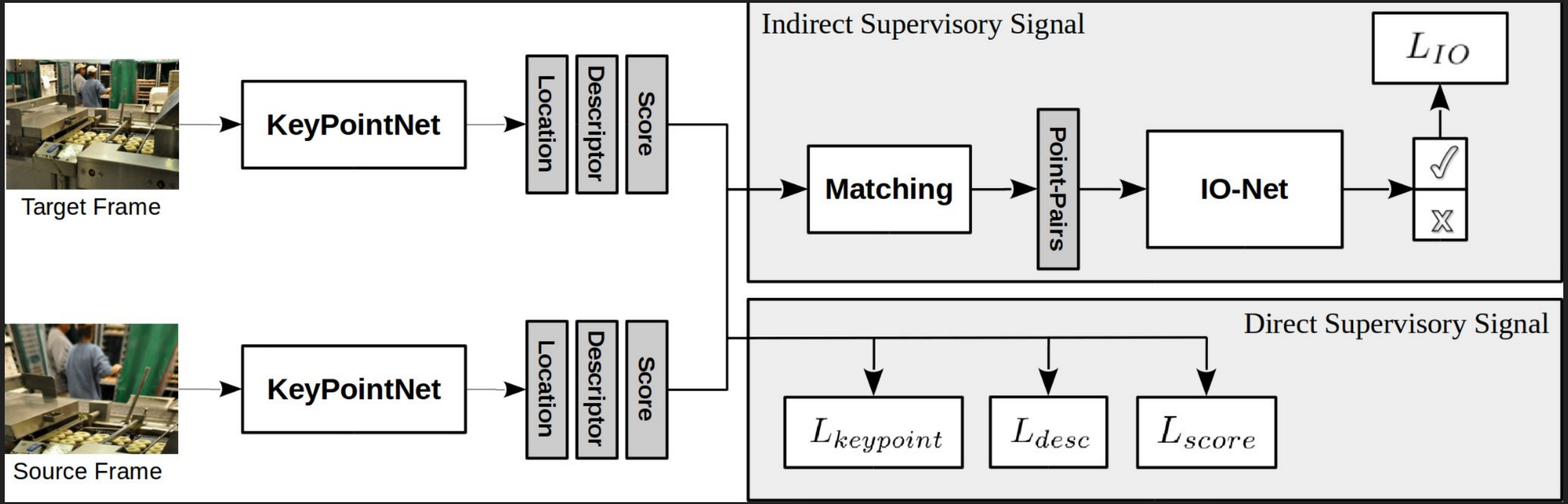
- Initial keypoint feature descriptors, coordinates, and scores from KP2D
- Filter out keypoints based on:
  - Score threshold
  - Descriptor-distance threshold
  - Distance on epipolar line
- Reproject keypoints left to 3D using depth map estimated from target image
- Compute pose using Perspective-n-Point algorithm with 2D-3D keypoint correspondences

1. Godard, C. et alii. (2019) Digging Into Self-Supervised Monocular Depth Estimation (ICCV)
2. J. Sturm and N. et alii. (2012) A Benchmark for the Evaluation of RGB-D SLAM Systems

PriorDepth – KP2D[1]

Extracts the keypoints, descriptors and the scores

Improve Outlier Rejection



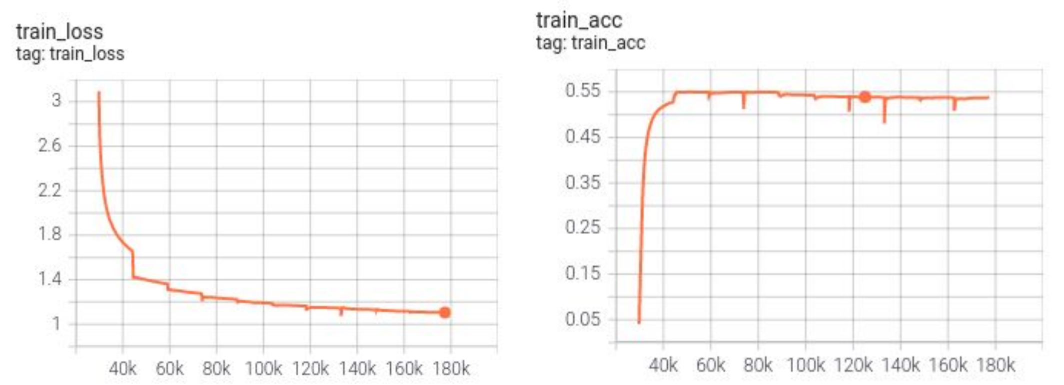
Keypoint Loss
Distance between the target keypoint and warped source keypoint

Descriptor Loss
Per pixel Triplet Loss on distance between the descriptors
+ve and -ve samples from keypoint correspondences between the images

Score Loss
Minimize the distance between scores of keypoint pairs + min./max. the average scores of keypoint pair

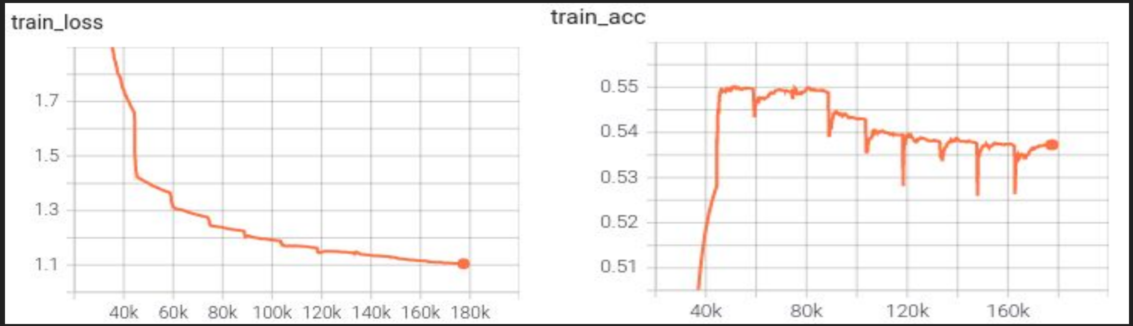
[1] Tang, Jiexiong, et al. “Neural Outlier Rejection for Self-Supervised Keypoint Learning.” ArXiv:1912.10615

Priordepth - KP2D



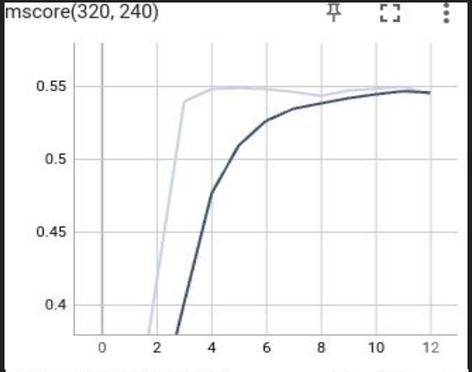
PriorDepth – KP2D

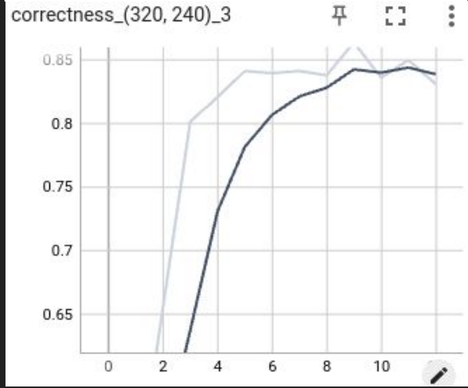
Training on COCO[2] 2017 (Train set)



Validation Metrics	Our Training (12 epochs)	Results from the paper (50 epochs)	Progress
C1	0.493	0.593	↑
C3	0.831	0.867	
C5	0.893	0.91	
Matching Score	0.544	0.546	
Repeatability	0.660	0.687	
Localization	0.913	0.892	↓

Validation on HPatches[3]





[2]Tsung-Yi Lin, et al. "Microsoft COCO: Common Objects in Context." (2015).
[3]Vassileios Balntas, et al. "HPatches: A benchmark and evaluation of handcrafted and learned local descriptors." (2017).

PriorDepth – Timeline

15.05 - 22.05

planning and research on individual network baselines

Focuses:
Halil: our baseline
Konstantin: Monodepth2[1]
Rahul: KeypointNet2D[2]

22.05 - 29.05

Simple testing on individual networks

AWS Instance environment setup

29.05 - 05.06

Offline training approaches

Preparing training datasets offline and on aws

05.06 - 12.06

AWS test training of networks

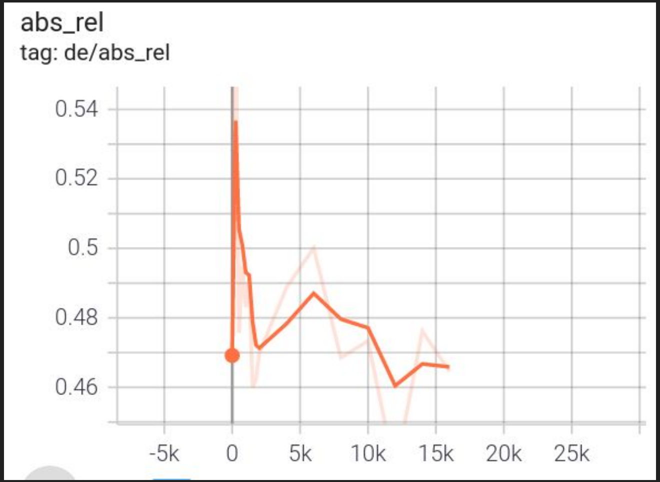
Building own Pipeline with trained networks

12.06 - 15.06

Debugging pipeline

1.Godard, C. et alii. (2019) Digging Into Self-Supervised Monocular Depth Estimation (ICCV)
2.Tang, J. et alii. (2019) Neural Outlier Rejection for Self-Supervised Keypoint Learning (ICLR)

PriorDepth – Monodepth2



Test Training Settings:

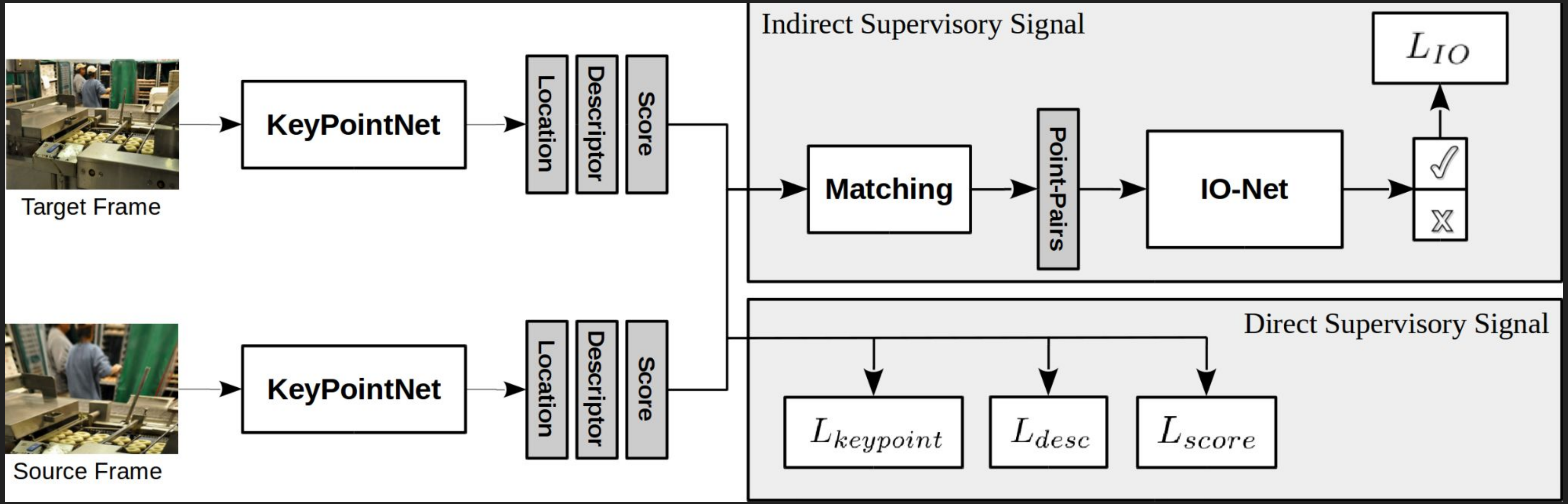
- 15 Epochs, adam optimizer, Batch size = 12, Lr = 1e-3
- Eigen Zhou split with around 44 000 images 10% validation 90% training
- From scratch

	Abs Rel	Sq Rel	RMSE	RMSE log
Test Model	0.4659	4.544	11.34	0.5796
Monodepth2 (after 20 epochs)	0.132	1.044	4.872	0.210

PriorDepth – KP2D[1]

Extracts the keypoints, descriptors and the scores

Improve Outlier Rejection



Keypoint Loss
Distance between the target keypoint and warped source keypoint

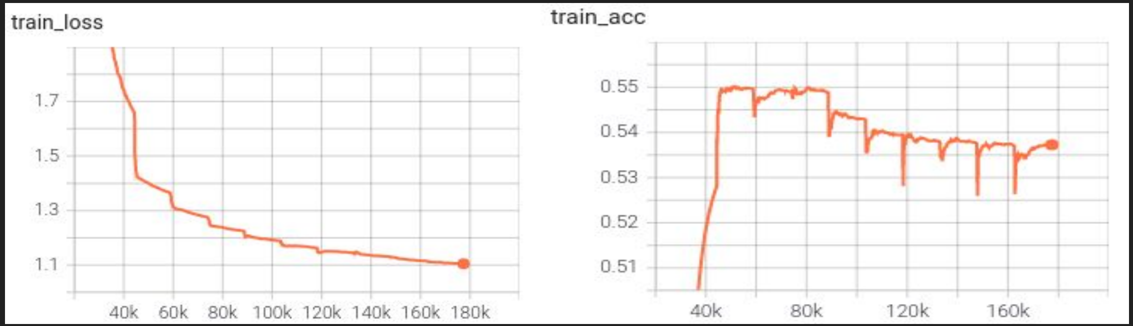
Descriptor Loss
Per pixel Triplet Loss on distance between the descriptors
+ve and -ve samples from keypoint correspondences between the images

Score Loss
Minimize the distance between scores of keypoint pairs + min./max. the average scores of keypoint pair

[1] Tang, Jiexiong, et al. “Neural Outlier Rejection for Self-Supervised Keypoint Learning.” ArXiv:1912.10615

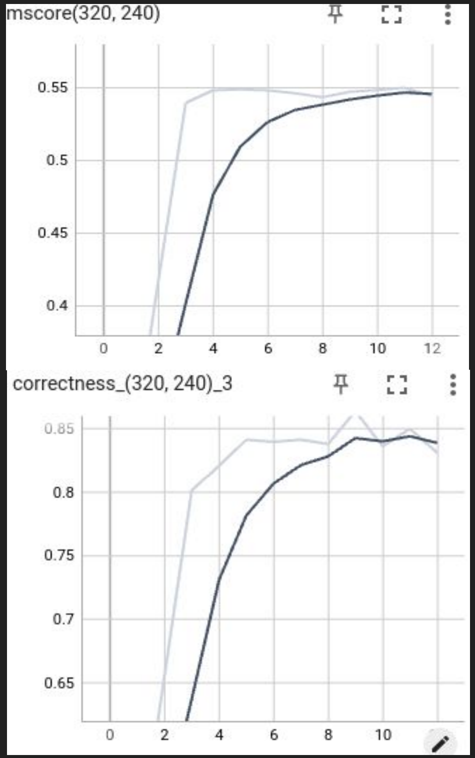
PriorDepth – KP2D

Training on COCO[2] 2017 (Train set)



Validation Metrics	Our Training (12 epochs)	Results from the paper (50 epochs)	Progress
C1	0.493	0.593	
C3	0.831	0.867	
C5	0.893	0.91	↑
Matching Score	0.544	0.546	
Repeatability	0.660	0.687	
Localization	0.913	0.892	↓

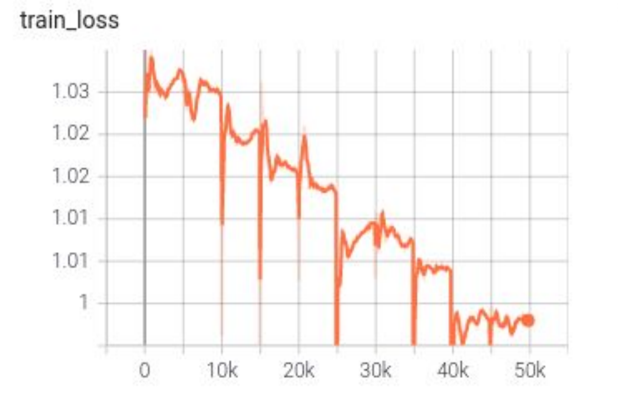
Validation on HPatches[3]

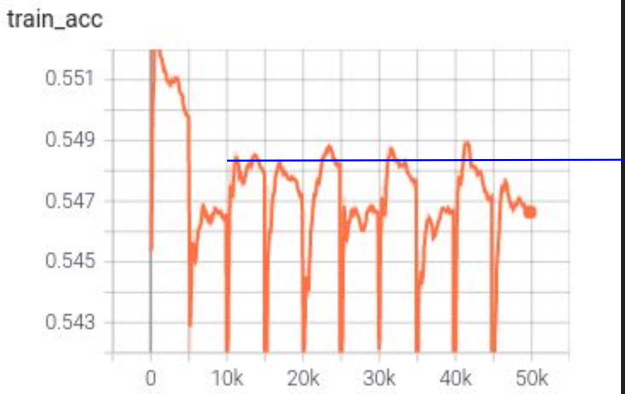


[2]Tsung-Yi Lin, et al. "Microsoft COCO: Common Objects in Context." (2015).
[3]Vassileios Balntas, et al. "HPatches: A benchmark and evaluation of handcrafted and learned local descriptors." (2017).

PriorDepth – KP2D

Pretraining on KITTI[4] - Eigen Zhou split





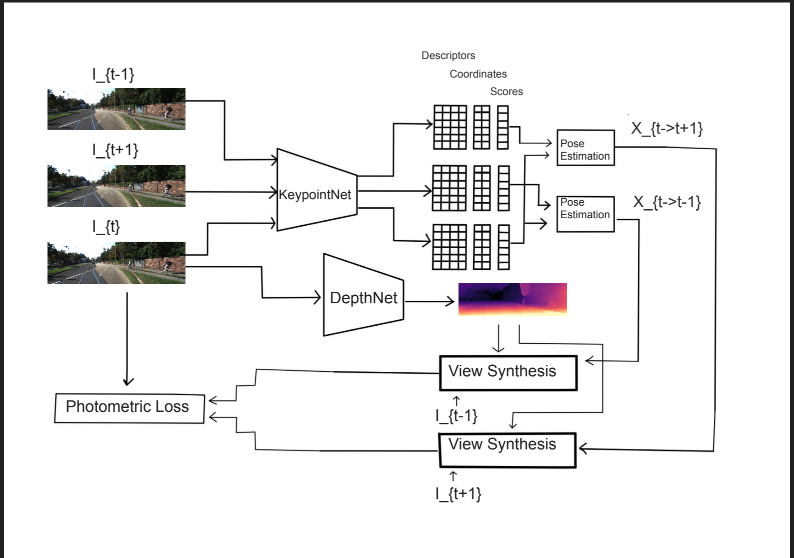
Help the network navigate the domain shift with pre-training

Use model that performs well on KITTI to plug into KP3D baseline model and freeze the network.

Figure - Visualisation of the matched keypoints on KITTI from KeypointNet

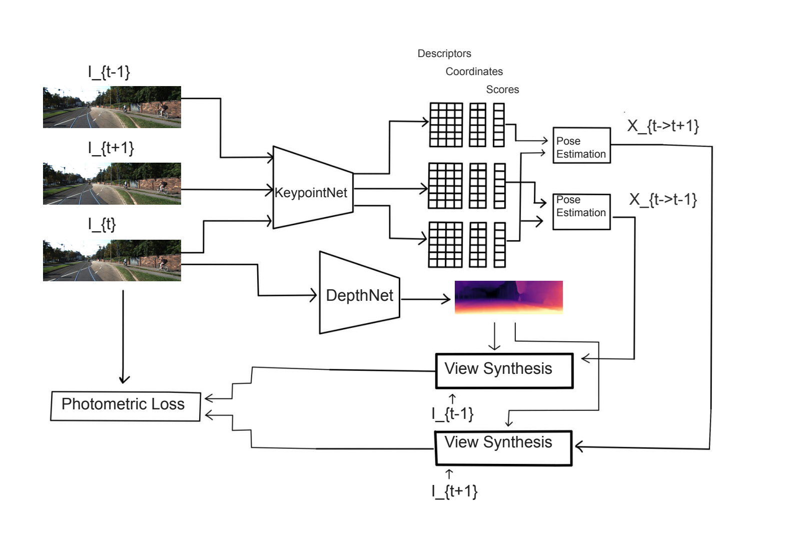
[4] Geiger, A., et al. “Vision Meets Robotics: The KITTI Dataset.” The International Journal of Robotics Research, vol. 32, no. 11, Sept. 2013

PriorDepth – KP3D Baseline

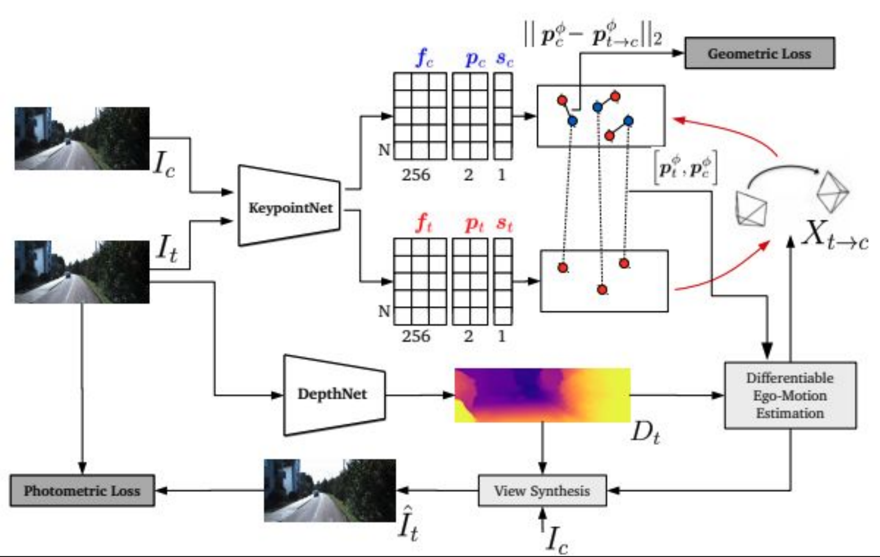


Our Baseline

PriorDepth – KP3D Baseline



Our Baseline

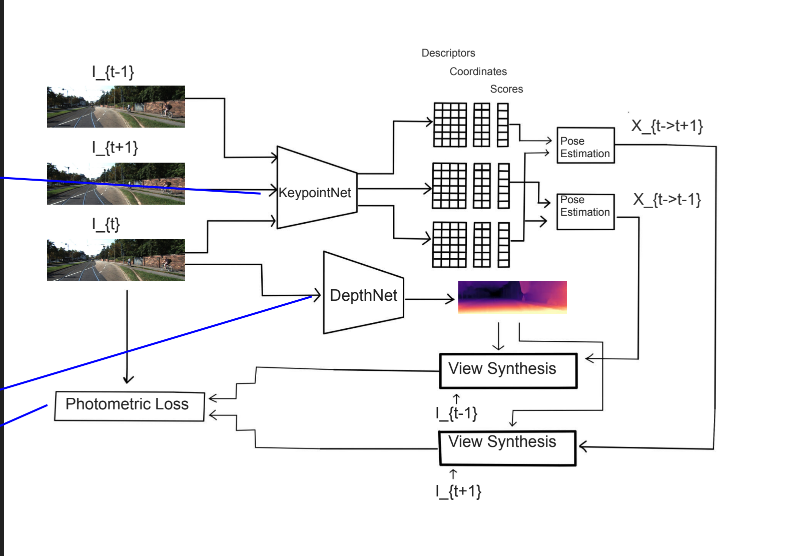


KP3D

PriorDepth – KP3D Baseline

From KP2D

Shared Encoder with Output Heads
Pre-trained on COCO, fine-tuned in KITTI
Freezed during training of KP3D



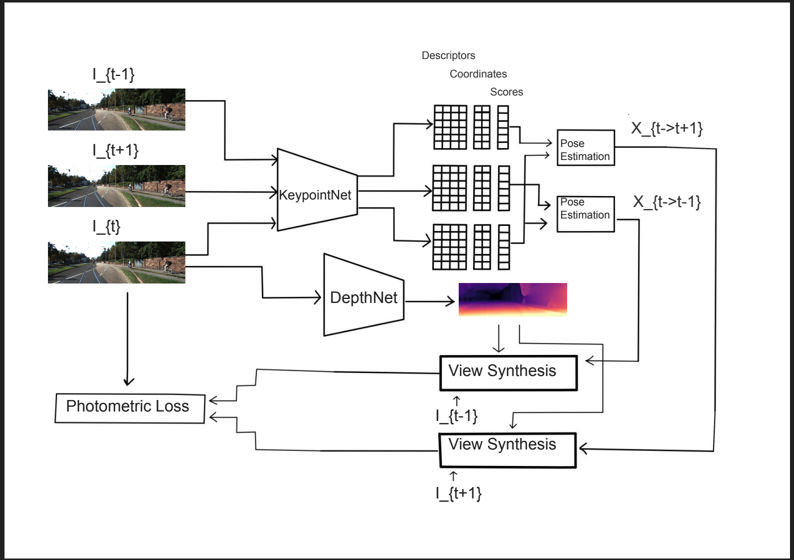
From MonoDepth2

Only Depth Encoder-Decoder is loaded
Pretrained on ImageNet, trained on KITTI
Photometric and Smooth Loss are utilized as depth losses

Build a KP3D based network

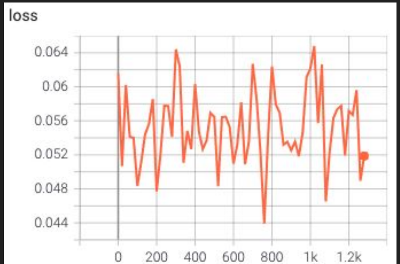
PriorDepth – KP3D Baseline

- Our current pipeline:
  - Input: Current and adjacent images where current is our target, adjacent images are contexts
  - Inverse depth estimation on Target image
  - Keypoint estimation on both target and context images
  - Pose estimation from target image to context images
  - Depth estimation from inverse depth map
  - View Synthesis utilizing depth maps, estimated poses, and context images
  - Photometric Loss calculation

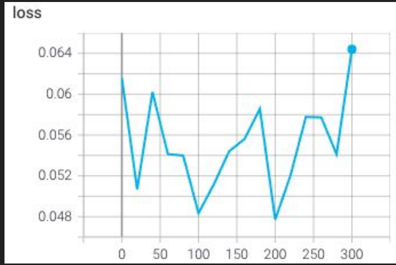


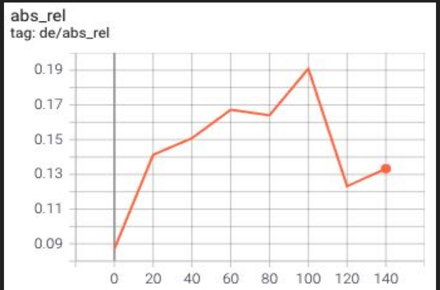
PriorDepth – KP3D Baseline

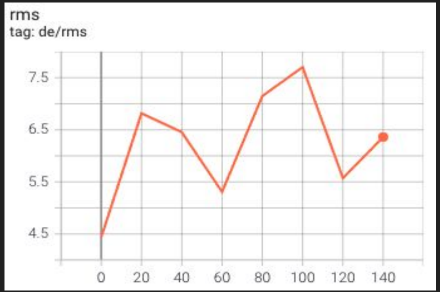
- Training Curves

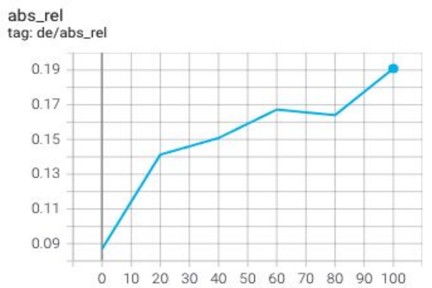


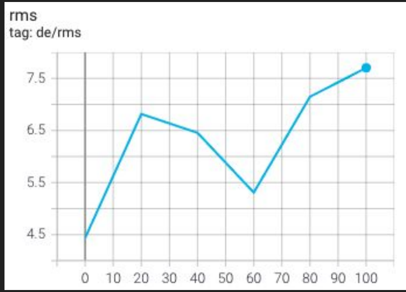
- Validation Curves







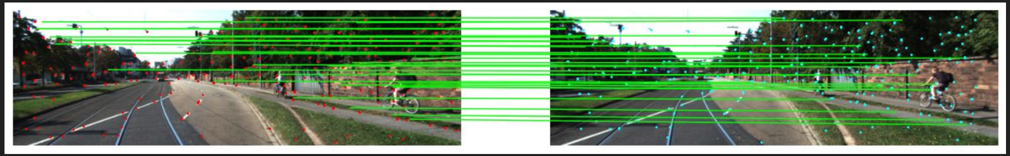




PriorDepth – KP3D Baseline

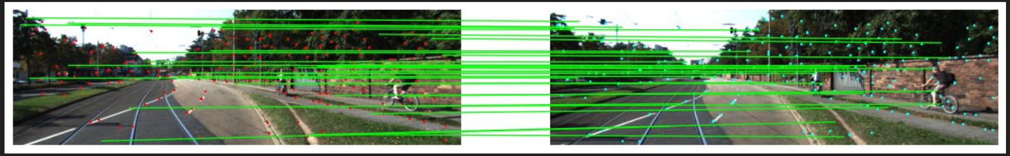
- Example visualizations

t-1



t

t+1



PriorDepth – Future Work

- Debug the network to be sure 100% everything is correct
- Additional visualizations for instance trajectory over time and warped images with estimated pose
- Additional evaluation metrics in addition to calculated training and depth losses
    - For example: Pose & Depth accuracy
- Training KP2D with MonoDepth2 together
    - Will implement Keypoint Loss
- Training on an indoor dataset to show applicability in various conditions such as camera motion
    - Camera in KITTI is almost stay still

- We may work on additional tasks as well, we will discuss :)

PriorDepth – Problem Statement

- **Problem**: Robust depth map and pose estimation using keypoints in self-supervised training
- **Solutions:**
  - MonoDepth2
    - Drawback: Pose Network does not work well in various scenes
  - KP2D
    - Drawback: Only using Homography Augmentations for 2D Warping
  - KP3D
    - Drawback: Does not use sparse triangulation for depth loss

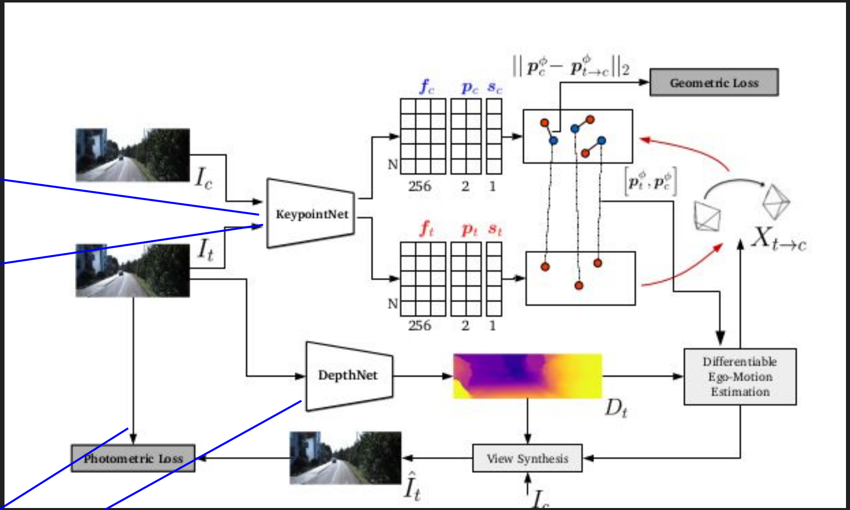
PriorDepth – Setting up the Model

KP2D based

3D warping for Keypoint Detection with estimated depth map

Joint Optimization of Depth Map and Keypoints

MonoDepth2 based



Sparse depth triangulation for depth reconstruction loss

Build a KP3D based network

PriorDepth – Related Work

- Related Work and existing Solutions

Task: Depth Estimation

Task: Keypoint Matching

Task: Ego-Motion Estimation

StereoBM based on Epipolar Geometry

SIFT/ORB based features

Non-Learning

MonoDepth2 [1]

KeyPoint2D [2]

KeyPoint3D [3]

Learning-based

[1] Clément Godard; Digging Into Self-Supervised Monocular Depth Estimation, ICCV 2018
[2] Jiexiong Tang; Neural Outlier Rejection for self-supervised keypoint learning, 2020
[3] Jiexiong Tang; Self-Supervised 3D Keypoint Learning for Ego-motion Estimation, 2020

Project PriorDepth – Conclusion

- ❏ Our task: Robust and accurate Depth and Pose Estimation

- ❏ What will be the (live) demo / prototype you want to show?

- ❏ We want to show improved depth and ego-motion estimations

Project X – Overview

- ❑ What is the general idea of the project?
- ❑ How can it be summarized?
- ❑ Think of TL;DR style

Project X – Motivation

- ❏ Why is it relevant / interesting?
- ❏ Where can it be used?
- ❏ Who benefits from it?
- ❏ What do you expect to learn?

Project X – Problem

- ❑ Summarize the problem
- ❑ Do solutions already exist?
- ❑ What is your method / strategy to solve it?
- ❑ Emphasize on why your method is suitable for it / what obstacles you see
- ❑ Can it be split in sub-problems?

Project X – Initial Plan

- ❏ Who is responsible for what?
- ❏ When do you plan to be ready with X1, X2, …?
- ❏ Plan more detailed until Project Update Presentations [15.06.2021]
- ❏ What are the to-dos afterwards until Final Workshop [15.07.2021]?
- ❏ What will be the (live) demo / prototype you want to show?