KI DELTA LEARNING

Scalable AI for Automated Driving

Final Event | March 09, 2023

A Benchmark and a Baseline for Robust Multi-view Depth Estimation

Philipp Schröppel

Outline

1. Introduction

   task description, related work, motivation

2. Robust Multi-view Depth Benchmark

   objective, test sets, evaluation settings, results

3. Robust MVD Baseline Model

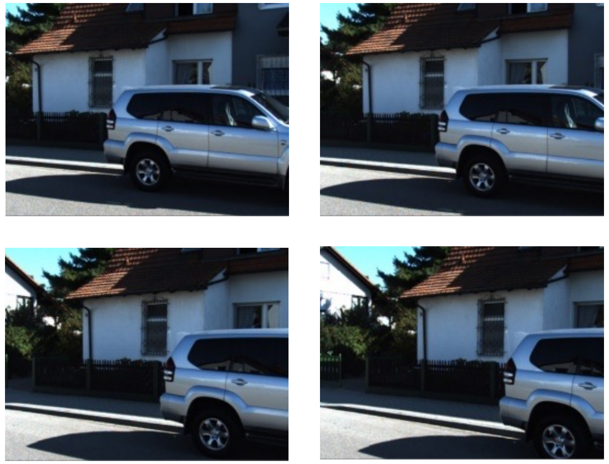
   overview, qualitative results

4. robustmvd Framework
5. Summary

Introduction
Multi-view Depth Estimation

Keyview V_0

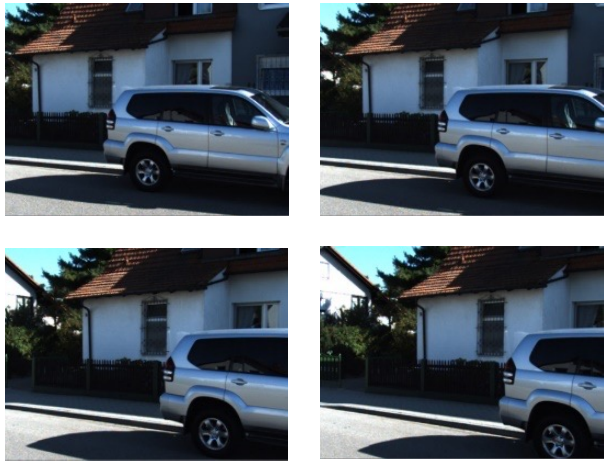
Source views $V_1,..,V_k$

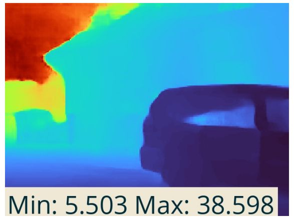


Introduction
Multi-view Depth Estimation

Keyview V_0

Source views $V_1, .., V_k$



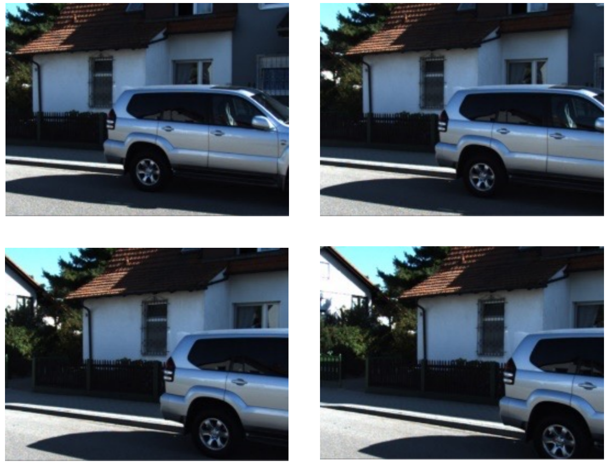


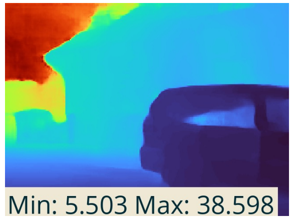
Depth map for the keyview

Introduction
Multi-view Depth Estimation

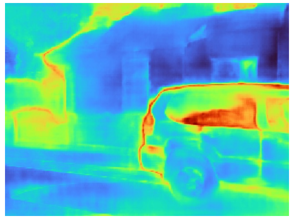
Keyview V_0

Source views $V_1,..,V_k$





Depth map for the keyview



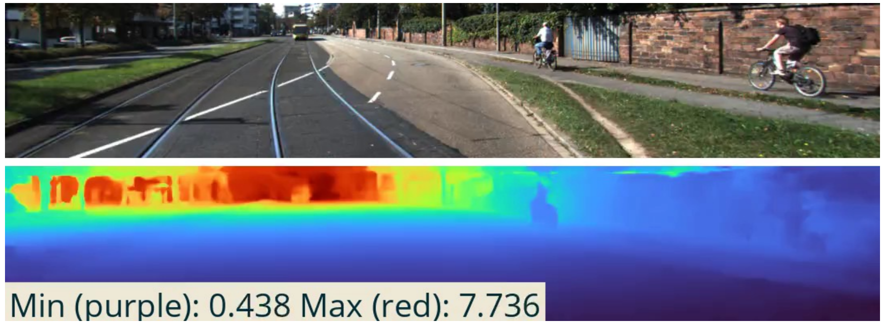
Depth Uncertainty

Introduction
Depth-from-video and Multi-view Stereo

Introduction
Depth-from-video and Multi-view Stereo

Depth-from-video:

- sequential input views
- **inputs**: images, intrinsics
- **evaluation**: align and compare predicted and ground truth depth maps

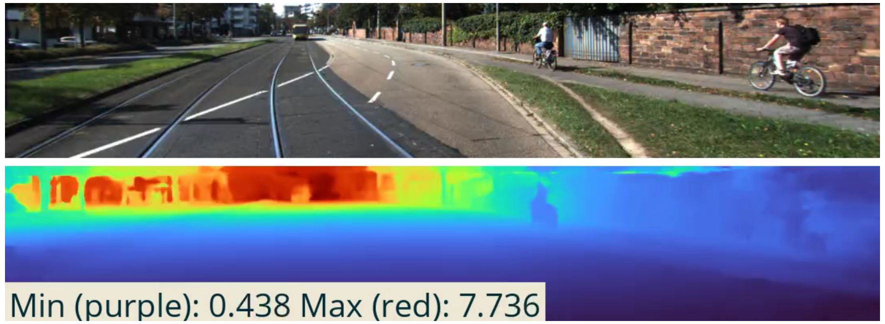


[DeepV2D, Teed and Deng]

Introduction
Depth-from-video and Multi-view Stereo

Depth-from-video:

- sequential input views
- **inputs**: images, intrinsics
- **evaluation**: align and compare predicted and ground truth depth maps



[DeepV2D, Teed and Deng]

Multi-view Stereo:

- unstructured input views
- **inputs**: images, intrinsics, poses, depth range
- **evaluation**: fuse depth maps and compare predicted and ground truth pointclouds

[Vis-MVSNet, Zhang et al.]

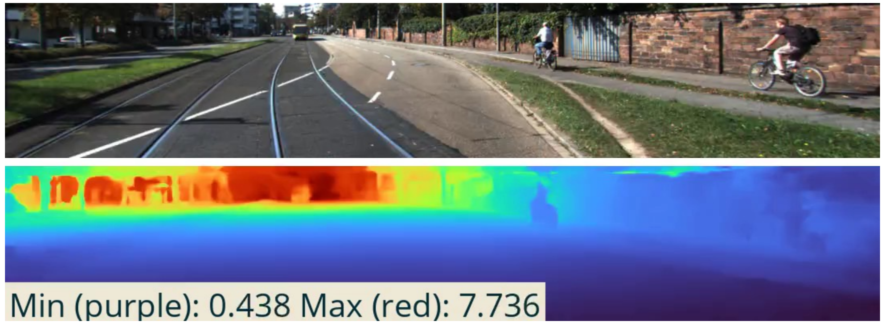
Introduction
Depth-from-video and Multi-view Stereo

Depth-from-video:

- sequential input views
- **inputs**: images, intrinsics
- **evaluation**: align and compare predicted and ground truth depth maps

Multi-view Stereo:

- unstructured input views
- **inputs**: images, intrinsics, poses, depth range
- **evaluation**: fuse depth maps and compare predicted and ground truth pointclouds



[DeepV2D, Teed and Deng]

[Vis-MVSNet, Zhang et al.]

→ depth is estimated from multiple input views

Introduction
Motivation for the Benchmark

Depth estimation is useful for many practical applications: autonomous driving, augmented reality, robotics, …

→ failures can be fatal

→ depth estimation should function robustly in an **open-world setting** with potentially unseen objects

→ multi-view depth estimation can derive depth from the **motion parallax** → generic principle that **should enable generalization**

→ however: training and testing is often done on similar data

Objective: evaluate multi-view depth estimation in an open-world setting

Robust Multi-view Depth Benchmark

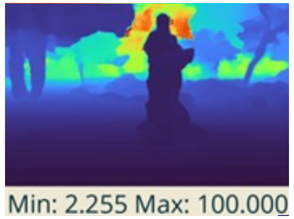
Benchmark multi-view depth estimation models regarding **robust application on arbitrary real-world data.**

Robust Multi-view Depth Benchmark

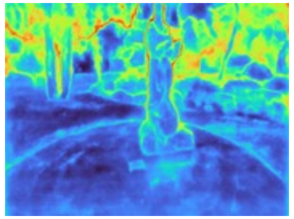
Benchmark multi-view depth estimation models regarding **robust application on arbitrary real-world data.**

Keyview V_0

Source views $V_1, .., V_k$



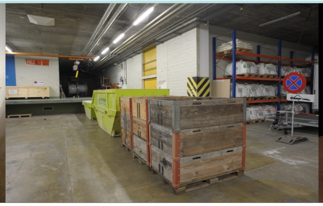
Depth map for the keyview



Depth Uncertainty

Robust Multi-view Depth Benchmark
Test Sets

Comprises test sets based on **diverse existing datasets:**

	KITTI	ScanNet	ETH3D	DTU	T & T
					
Domain	Driving	Indoor	In- & outdoor	Tabletop	In- & outdoor
Structure	Video	Video	None	None	None
Scene scale	2 – 85m	0.2 – 0.9m	0.3 – 60m	0.4 – 1.2m	1.1 – 42m
# samples	93	200	104	110	69

→ training data intentionally left undefined

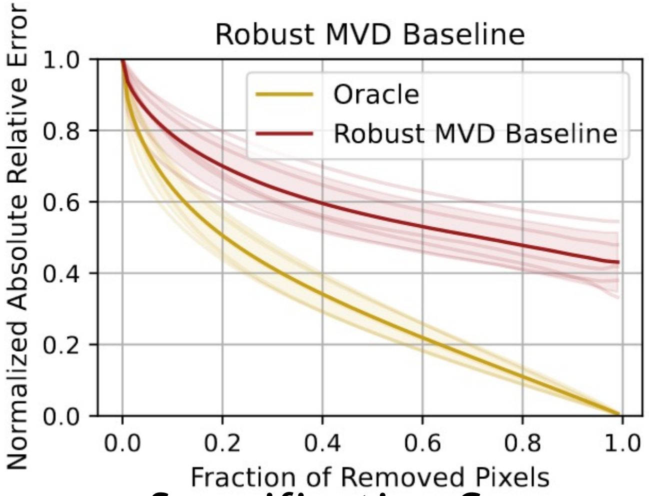
→ evaluation in a zero-shot cross-dataset fashion

Robust Multi-view Depth Benchmark
Settings and Metrics

- Features different evaluation settings:
  - input modalities: images, intrinsics, poses, depth range
  - optional alignment between predicted and GT depth maps

- Depth estimation metrics:
  - Absolute Relative Error (rel)
  - Inlier Ratio with a Threshold of 1.03 (τ)

- Uncertainty metrics:
  - Sparsification Error Curves
  - Area Under Sparsification Error (AUSE)



Sparsification Curves

Robust Multi-view Depth Benchmark
Results

Approach	GT Poses	GT Range	Align	KITTI rel ↓	KITTI τ ↑	ScanNet rel ↓	ScanNet τ ↑	ETH3D rel ↓	ETH3D τ ↑	DTU rel ↓	DTU τ ↑	T&T rel ↓	T&T τ ↑	Average rel ↓	Average τ ↑	Average time [s] ↓
COLMAP	✓	✗	✗	**12.0**	**58.2**	**14.6**	**34.2**	**16.4**	**55.1**	**0.7**	**96.5**	**2.7**	**95.0**	**9.3**	**67.8**	≈ 3 min
COLMAP Dense	✓	✗	✗	26.9	52.7	38.0	22.5	89.8	23.2	20.8	69.3	25.7	76.4	40.2	48.8	≈ 3 min
DeMoN	✗	✗	$\|\|t\|\|$	15.5	15.2	**12.0**	**21.0**	17.4	15.4	21.8	16.6	13.0	23.2	16.0	18.3	**0.08**
DeepV2D KITTI	✗	✗	med	**(3.1)**	**(74.9)**	23.7	11.1	27.1	10.1	24.8	8.1	34.1	9.1	22.6	22.7	2.07
DeepV2D ScanNet	✗	✗	med	**10.0**	**36.2**	(4.4)	(54.8)	**11.8**	**29.3**	**7.7**	**33.0**	**8.9**	**46.4**	**8.6**	**39.9**	3.57
MVSNet	✓	✓	✗	22.7	36.1	24.6	20.4	35.4	31.4	**(1.8)**	(86.0)	8.3	73.0	18.6	49.4	0.07
MVSNet Inv. Depth	✓	✓	✗	18.6	30.7	22.7	20.9	21.6	35.6	**(1.8)**	(86.7)	6.5	74.6	14.2	49.7	0.32
CVP-MVSNet	✓	✓	✗	156.7	2.2	137.1	15.9	156.4	13.6	(4.0)	(68.4)	24.7	52.9	95.8	30.6	0.49
Vis-MVSNet	✓	✓	✗	**9.5**	**55.4**	8.9	33.5	**10.8**	**43.3**	**(1.8)**	**(87.4)**	**4.1**	**87.2**	**7.0**	**61.4**	0.70
PatchmatchNet	✓	✓	✗	10.8	45.8	**8.5**	**35.3**	19.1	34.8	(2.1)	(82.8)	4.8	82.9	9.1	56.3	0.28
Fast-MVSNet	✓	✓	✗	14.4	37.1	17.0	24.6	25.2	32.0	(2.5)	(81.8)	8.3	68.6	13.5	48.8	0.30
MVS2D ScanNet	✓	✓	✗	21.2	8.7	(27.2)	(5.3)	27.4	4.8	17.2	9.8	29.2	4.4	24.4	6.6	**0.04**
MVS2D DTU	✓	✓	✗	226.6	0.7	32.3	11.1	99.0	11.6	(3.6)	(64.2)	25.8	28.0	77.5	23.1	0.05

bold: best | (parentheses): trained on same domain

Robust Multi-view Depth Benchmark
Results

Approach	GT Poses	GT Range	Align	KITTI rel ↓	KITTI τ ↑	ScanNet rel ↓	ScanNet τ ↑	ETH3D rel ↓	ETH3D τ ↑	DTU rel ↓	DTU τ ↑	T&T rel ↓	T&T τ ↑	Average rel ↓	Average τ ↑	Average time [s] ↓
COLMAP	✓	✗	✗	**12.0**	**58.2**	**14.6**	**34.2**	**16.4**	**55.1**	**0.7**	**96.5**	**2.7**	**95.0**	**9.3**	**67.8**	≈ 3 min
COLMAP Dense	✓	✗	✗	26.9	52.7	38.0	22.5	89.8	23.2	20.8	69.3	25.7	76.4	40.2	48.8	≈ 3 min
DeMoN	✗	✗	$\|t\|$	15.5	15.2	**12.0**	**21.0**	17.4	15.4	21.8	16.6	13.0	23.2	16.0	18.3	**0.08**
DeepV2D KITTI	✗	✗	med	**(3.1)**	**(74.9)**	23.7	11.1	27.1	10.1	24.8	8.1	34.1	9.1	22.6	22.7	2.07
DeepV2D ScanNet	✗	✗	med	**10.0**	**36.2**	**(4.4)**	**(54.8)**	**11.8**	**29.3**	**7.7**	**33.0**	**8.9**	**46.4**	**8.6**	**39.9**	3.57
MVSNet	✓	✓	✗	22.7	36.1	24.6	20.4	35.4	31.4	**(1.8)**	(86.0)	8.3	73.0	18.6	49.4	0.07
MVSNet Inv. Depth	✓	✓	✗	18.6	30.7	22.7	20.9	21.6	35.6	**(1.8)**	(86.7)	6.5	74.6	14.2	49.7	0.32
CVP-MVSNet	✓	✓	✗	156.7	2.2	137.1	15.9	156.4	13.6	(4.0)	(68.4)	24.7	52.9	95.8	30.6	0.49
Vis-MVSNet	✓	✓	✗	**9.5**	**55.4**	8.9	33.5	**10.8**	**43.3**	**(1.8)**	**(87.4)**	**4.1**	**87.2**	**7.0**	**61.4**	0.70
PatchmatchNet	✓	✓	✗	10.8	45.8	**8.5**	**35.3**	19.1	34.8	(2.1)	(82.8)	4.8	82.9	9.1	56.3	0.28
Fast-MVSNet	✓	✓	✗	14.4	37.1	17.0	24.6	25.2	32.0	(2.5)	(81.8)	8.3	68.6	13.5	48.8	0.30
MVS2D ScanNet	✓	✓	✗	21.2	8.7	(27.2)	(5.3)	27.4	4.8	17.2	9.8	29.2	4.4	24.4	6.6	**0.04**
MVS2D DTU	✓	✓	✗	226.6	0.7	32.3	11.1	99.0	11.6	(3.6)	(64.2)	25.8	28.0	77.5	23.1	0.05

bold: best | (parentheses): trained on same domain

Robust Multi-view Depth Benchmark

Results

Approach	GT Poses	GT Range	Align	KITTI rel ↓	KITTI τ ↑	ScanNet rel ↓	ScanNet τ ↑	ETH3D rel ↓	ETH3D τ ↑	DTU rel ↓	DTU τ ↑	T&T rel ↓	T&T τ ↑	Average rel ↓	Average τ ↑	Average time [s] ↓
COLMAP	✓	✗	✗	**12.0**	**58.2**	**14.6**	**34.2**	**16.4**	**55.1**	**0.7**	**96.5**	**2.7**	**95.0**	**9.3**	**67.8**	≈ 3 min
COLMAP Dense	✓	✗	✗	26.9	52.7	38.0	22.5	89.8	23.2	20.8	69.3	25.7	76.4	40.2	48.8	≈ 3 min
DeMoN	✗	✗	$\|\|t\|\|$	15.5	15.2	**12.0**	**21.0**	17.4	15.4	21.8	16.6	13.0	23.2	16.0	18.3	**0.08**
DeepV2D KITTI	✗	✗	med	**(3.1)**	**(74.9)**	23.7	11.1	27.1	10.1	24.8	8.1	34.1	9.1	22.6	22.7	2.07
DeepV2D ScanNet	✗	✗	med	**10.0**	**36.2**	(4.4)	(54.8)	**11.8**	**29.3**	**7.7**	**33.0**	**8.9**	**46.4**	**8.6**	**39.9**	3.57
MVSNet	✓	✓	✗	22.7	36.1	24.6	20.4	35.4	31.4	**(1.8)**	(86.0)	8.3	73.0	18.6	49.4	0.07
MVSNet Inv. Depth	✓	✓	✗	18.6	30.7	22.7	20.9	21.6	35.6	**(1.8)**	(86.7)	6.5	74.6	14.2	49.7	0.32
CVP-MVSNet	✓	✓	✗	156.7	2.2	137.1	15.9	156.4	13.6	(4.0)	(68.4)	24.7	52.9	95.8	30.6	0.49
Vis-MVSNet	✓	✓	✗	**9.5**	**55.4**	8.9	33.5	**10.8**	**43.3**	**(1.8)**	**(87.4)**	**4.1**	**87.2**	**7.0**	**61.4**	0.70
PatchmatchNet	✓	✓	✗	10.8	45.8	**8.5**	**35.3**	19.1	34.8	(2.1)	(82.8)	4.8	82.9	9.1	56.3	0.28
Fast-MVSNet	✓	✓	✗	14.4	37.1	17.0	24.6	25.2	32.0	(2.5)	(81.8)	8.3	68.6	13.5	48.8	0.30
MVS2D ScanNet	✓	✓	✗	21.2	8.7	(27.2)	(5.3)	27.4	4.8	17.2	9.8	29.2	4.4	24.4	6.6	**0.04**
MVS2D DTU	✓	✓	✗	226.6	0.7	32.3	11.1	99.0	11.6	(3.6)	(64.2)	25.8	28.0	77.5	23.1	0.05

bold: best | (parentheses): trained on same domain

Robust Multi-view Depth Benchmark
Results

Approach	GT Poses	GT Range	Align	KITTI rel ↓	KITTI τ ↑	ScanNet rel ↓	ScanNet τ ↑	ETH3D rel ↓	ETH3D τ ↑	DTU rel ↓	DTU τ ↑	T&T rel ↓	T&T τ ↑	Average rel ↓	Average τ ↑	Average time [s] ↓
DeMoN	✓	✗	✗	16.7	13.4	75.0	0.0	19.0	16.2	23.7	11.5	17.6	18.3	30.4	11.9	0.08
DeepTAM	✓	✗	✗	68.7	0.4	(6.7)	(39.7)	20.4	19.8	58.0	9.1	40.0	12.9	38.8	16.4	0.85
DeepV2D KITTI	✓	✗	✗	(20.4)	(16.3)	25.8	8.1	30.1	9.4	24.6	8.2	38.5	9.6	27.9	10.3	1.43
DeepV2D ScanNet	✓	✗	✗	61.9	5.2	**(3.8)**	**(60.2)**	18.7	28.7	9.2	27.4	33.5	38.0	25.4	31.9	2.15
MVSNet	✓	✗	✗	14.0	35.8	1568.0	5.7	507.7	8.3	(4429.1)	(0.1)	118.2	50.7	1327.4	20.1	0.15
MVSNet Inv. Depth	✓	✗	✗	29.6	8.1	65.2	28.5	60.3	5.8	(28.7)	(48.9)	51.4	14.6	47.0	21.2	0.28
CVP-MVSNet	✓	✗	✗	158.2	1.2	2289.0	0.1	1735.3	1.2	(8314.0)	(0.0)	415.9	9.5	2582.5	2.4	0.50
Vis-MVSNet	✓	✗	✗	10.3	**54.4**	84.9	15.6	51.5	17.4	(374.2)	(1.7)	21.1	65.6	108.4	31.0	0.82
PatchmatchNet	✓	✗	✗	29.0	16.3	70.1	16.7	99.4	3.5	(82.6)	(5.6)	39.4	19.3	64.1	12.3	0.18
Fast-MVSNet	✓	✗	✗	12.1	37.4	287.1	9.4	131.2	9.6	(540.4)	(1.9)	33.9	47.2	200.9	21.1	0.35
MVS2D ScanNet	✓	✗	✗	73.4	0.0	(4.5)	(54.1)	30.7	14.4	5.0	57.9	56.4	11.1	34.0	27.5	**0.05**
MVS2D DTU	✓	✗	✗	93.3	0.0	51.5	1.6	78.0	0.0	**(1.6)**	**(92.3)**	87.5	0.0	62.4	18.8	0.06
Robust MVD Baseline	✓	✗	✗	**7.1**	41.9	**7.4**	**38.4**	**9.0**	**42.6**	**2.7**	**82.0**	**5.0**	**75.1**	**6.3**	**56.0**	0.06

bold: best | (parentheses): trained on same domain

Robust Multi-view Depth Benchmark
Results

Approach	GT Poses	GT Range	Align	KITTI		ScanNet		ETH3D		DTU		T&T		Average		
				rel $\downarrow$	$\tau\uparrow$	rel $\downarrow$	$\tau\uparrow$	rel $\downarrow$	$\tau\uparrow$	rel $\downarrow$	$\tau\uparrow$	rel $\downarrow$	$\tau\uparrow$	rel $\downarrow$	$\tau\uparrow$	time [s] $\downarrow$
DeMoN	✓	✗	✗	16.7	13.4	75.0	0.0	19.0	16.2	23.7	11.5	17.6	18.3	30.4	11.9	0.08
DeepTAM	✓	✗	✗	68.7	0.4	(6.7)	(39.7)	20.4	19.8	58.0	9.1	40.0	12.9	38.8	16.4	0.85
DeepV2D KITTI	✓	✗	✗	(20.4)	(16.3)	25.8	8.1	30.1	9.4	24.6	8.2	38.5	9.6	27.9	10.3	1.43
DeepV2D ScanNet	✓	✗	✗	61.9	5.2	**(3.8)**	**(60.2)**	18.7	28.7	9.2	27.4	33.5	38.0	25.4	31.9	2.15
MVSNet	✓	✗	✗	14.0	35.8	1568.0	5.7	507.7	8.3	(4429.1)	(0.1)	118.2	50.7	1327.4	20.1	0.15
MVSNet Inv. Depth	✓	✗	✗	29.6	8.1	65.2	28.5	60.3	5.8	(28.7)	(48.9)	51.4	14.6	47.0	21.2	0.28
CVP-MVSNet	✓	✗	✗	158.2	1.2	2289.0	0.1	1735.3	1.2	(8314.0)	(0.0)	415.9	9.5	2582.5	2.4	0.50
Vis-MVSNet	✓	✗	✗	10.3	**54.4**	84.9	15.6	51.5	17.4	(374.2)	(1.7)	21.1	65.6	108.4	31.0	0.82
PatchmatchNet	✓	✗	✗	29.0	16.3	70.1	16.7	99.4	3.5	(82.6)	(5.6)	39.4	19.3	64.1	12.3	0.18
Fast-MVSNet	✓	✗	✗	12.1	37.4	287.1	9.4	131.2	9.6	(540.4)	(1.9)	33.9	47.2	200.9	21.1	0.35
MVS2D ScanNet	✓	✗	✗	73.4	0.0	(4.5)	(54.1)	30.7	14.4	5.0	57.9	56.4	11.1	34.0	27.5	**0.05**
MVS2D DTU	✓	✗	✗	93.3	0.0	51.5	1.6	78.0	0.0	**(1.6)**	**(92.3)**	87.5	0.0	62.4	18.8	0.06
Robust MVD Baseline	✓	✗	✗	**7.1**	41.9	**7.4**	**38.4**	**9.0**	**42.6**	**2.7**	**82.0**	**5.0**	**75.1**	**6.3**	**56.0**	0.06

bold: best | (parentheses): trained on same domain

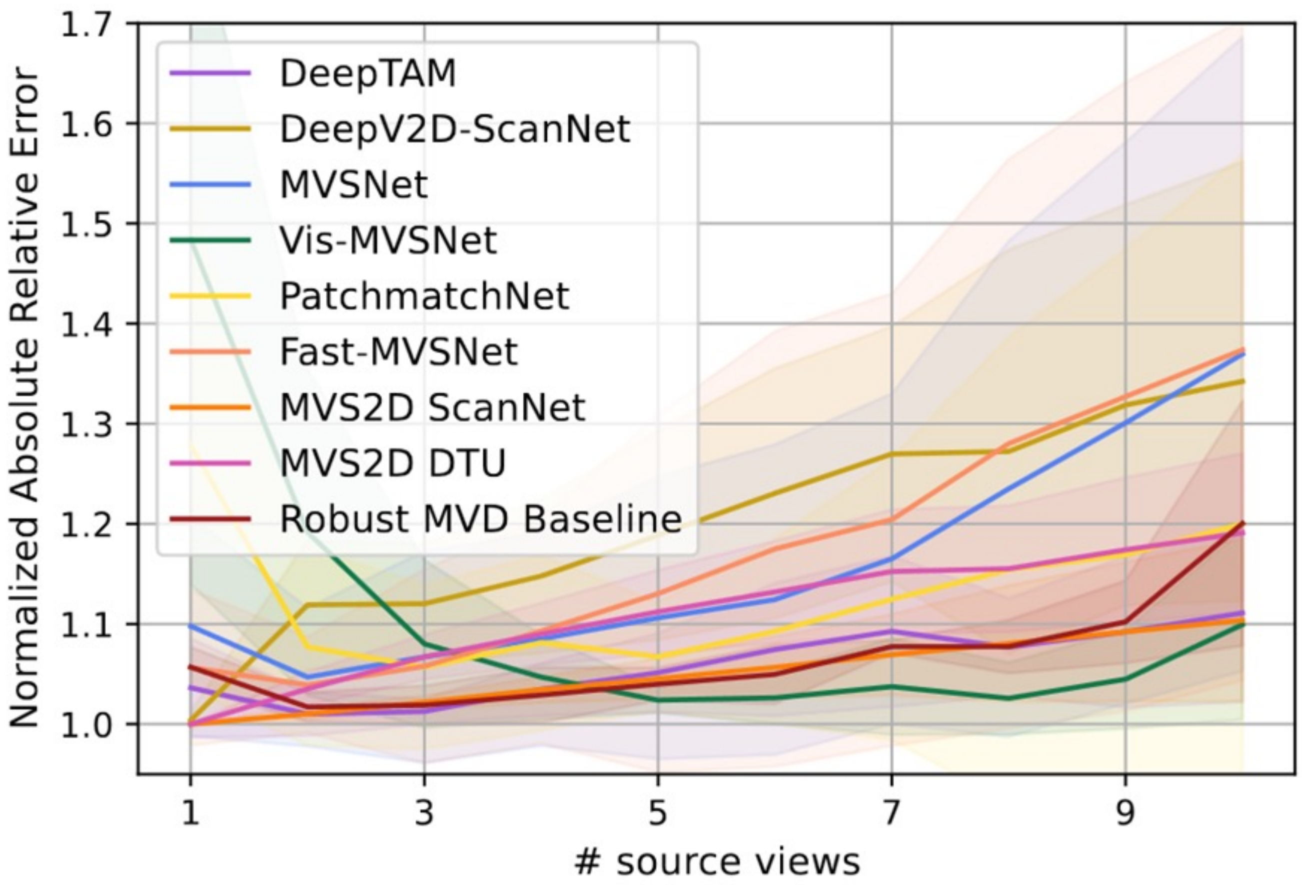
Robust Multi-view Depth Benchmark

Results

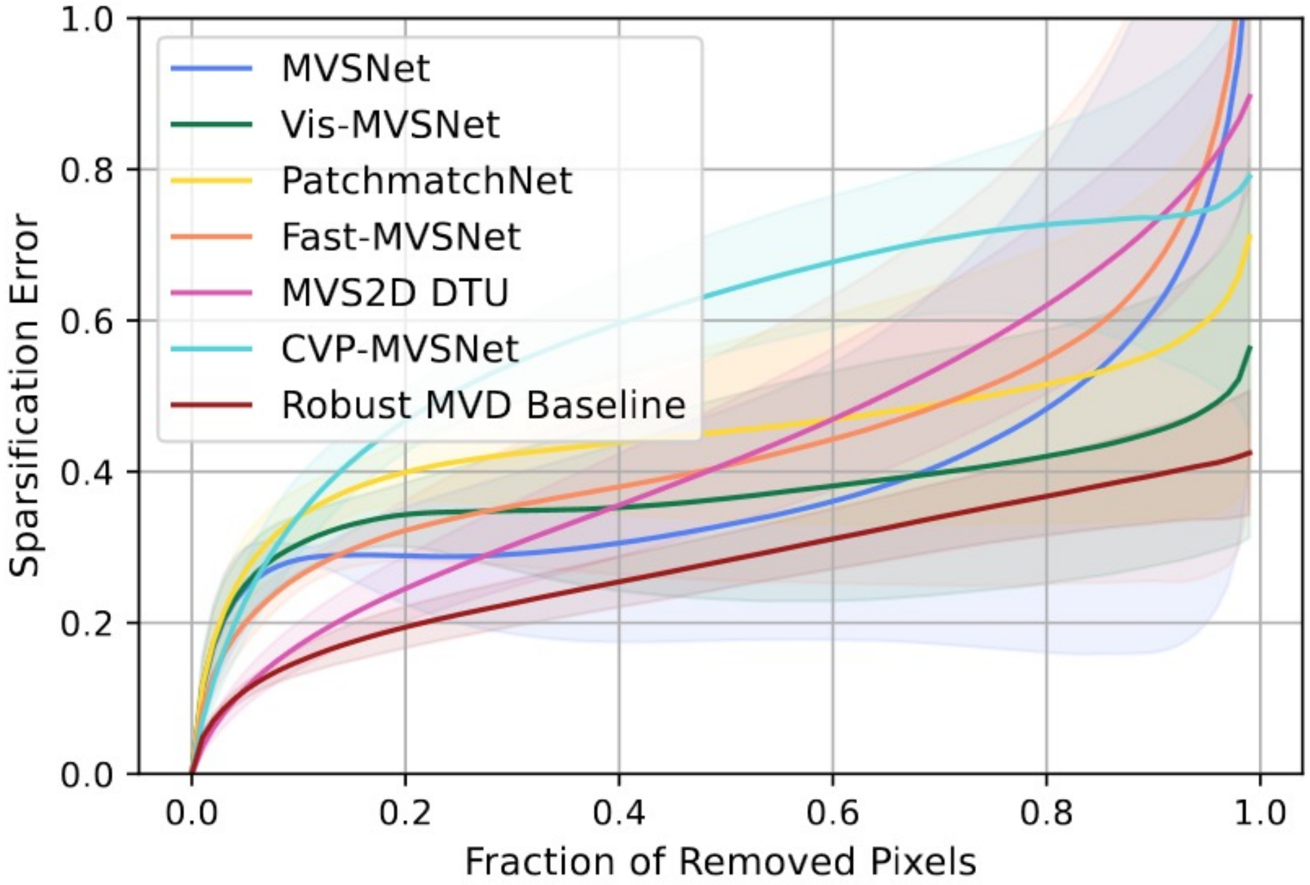
Approach	GT Poses	GT Range	Align	KITTI		ScanNet		ETH3D		DTU		T&T		Average		
				rel ↓	τ ↑	rel ↓	τ ↑	rel ↓	τ ↑	rel ↓	τ ↑	rel ↓	τ ↑	rel ↓	τ ↑	time [s] ↓
DeMoN	✓	✗	✗	16.7	13.4	75.0	0.0	19.0	16.2	23.7	11.5	17.6	18.3	30.4	11.9	0.08
DeepTAM	✓	✗	✗	68.7	0.4	(6.7)	(39.7)	20.4	19.8	58.0	9.1	40.0	12.9	38.8	16.4	0.85
DeepV2D KITTI	✓	✗	✗	(20.4)	(16.3)	25.8	8.1	30.1	9.4	24.6	8.2	38.5	9.6	27.9	10.3	1.43
DeepV2D ScanNet	✓	✗	✗	61.9	5.2	**(3.8)**	**(60.2)**	18.7	28.7	9.2	27.4	33.5	38.0	25.4	31.9	2.15
MVSNet	✓	✗	✗	14.0	35.8	1568.0	5.7	507.7	8.3	(4429.1)	(0.1)	118.2	50.7	1327.4	20.1	0.15
MVSNet Inv. Depth	✓	✗	✗	29.6	8.1	65.2	28.5	60.3	5.8	(28.7)	(48.9)	51.4	14.6	47.0	21.2	0.28
CVP-MVSNet	✓	✗	✗	158.2	1.2	2289.0	0.1	1735.3	1.2	(8314.0)	(0.0)	415.9	9.5	2582.5	2.4	0.50
Vis-MVSNet	✓	✗	✗	10.3	**54.4**	84.9	15.6	51.5	17.4	(374.2)	(1.7)	21.1	65.6	108.4	31.0	0.82
PatchmatchNet	✓	✗	✗	29.0	16.3	70.1	16.7	99.4	3.5	(82.6)	(5.6)	39.4	19.3	64.1	12.3	0.18
Fast-MVSNet	✓	✗	✗	12.1	37.4	287.1	9.4	131.2	9.6	(540.4)	(1.9)	33.9	47.2	200.9	21.1	0.35
MVS2D ScanNet	✓	✗	✗	73.4	0.0	(4.5)	(54.1)	30.7	14.4	5.0	57.9	56.4	11.1	34.0	27.5	**0.05**
MVS2D DTU	✓	✗	✗	93.3	0.0	51.5	1.6	78.0	0.0	**(1.6)**	**(92.3)**	87.5	0.0	62.4	18.8	0.06
Robust MVD Baseline	✓	✗	✗	**7.1**	41.9	**7.4**	**38.4**	**9.0**	**42.6**	**2.7**	**82.0**	**5.0**	**75.1**	**6.3**	**56.0**	0.06

bold: best | (parentheses): trained on same domain

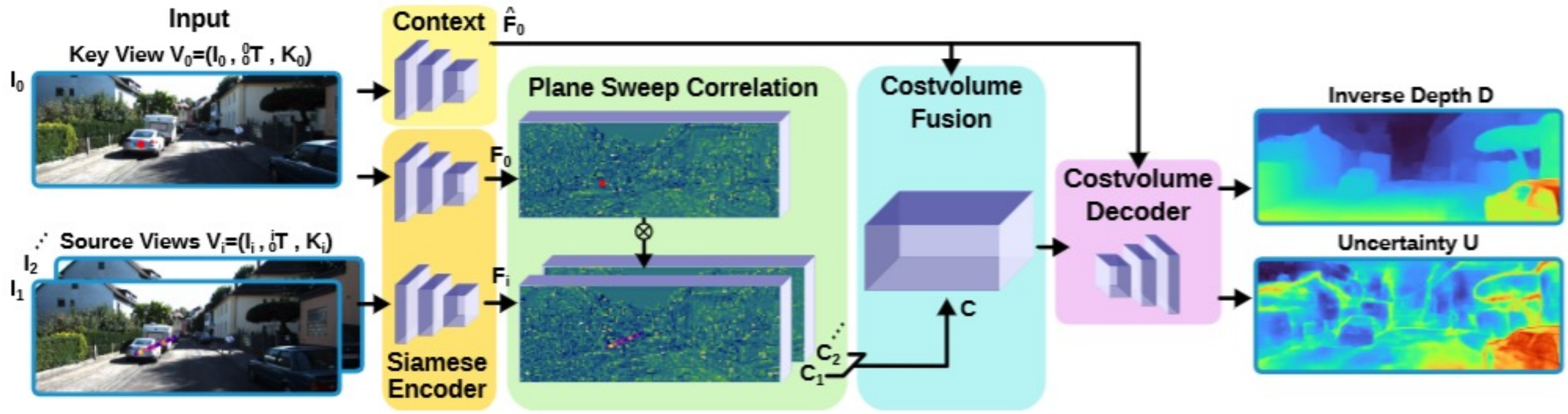
Robust Multi-view Depth Benchmark
Multi-view Fusion Results



Robust Multi-view Depth Benchmark
Uncertainty Estimation Results



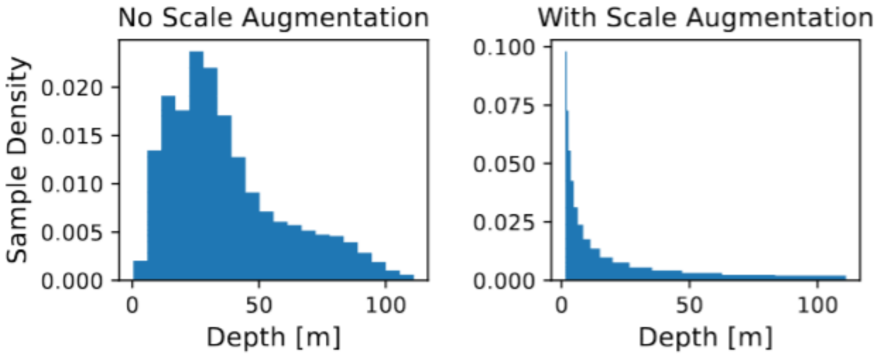
Robust MVD Baseline Model
Overview



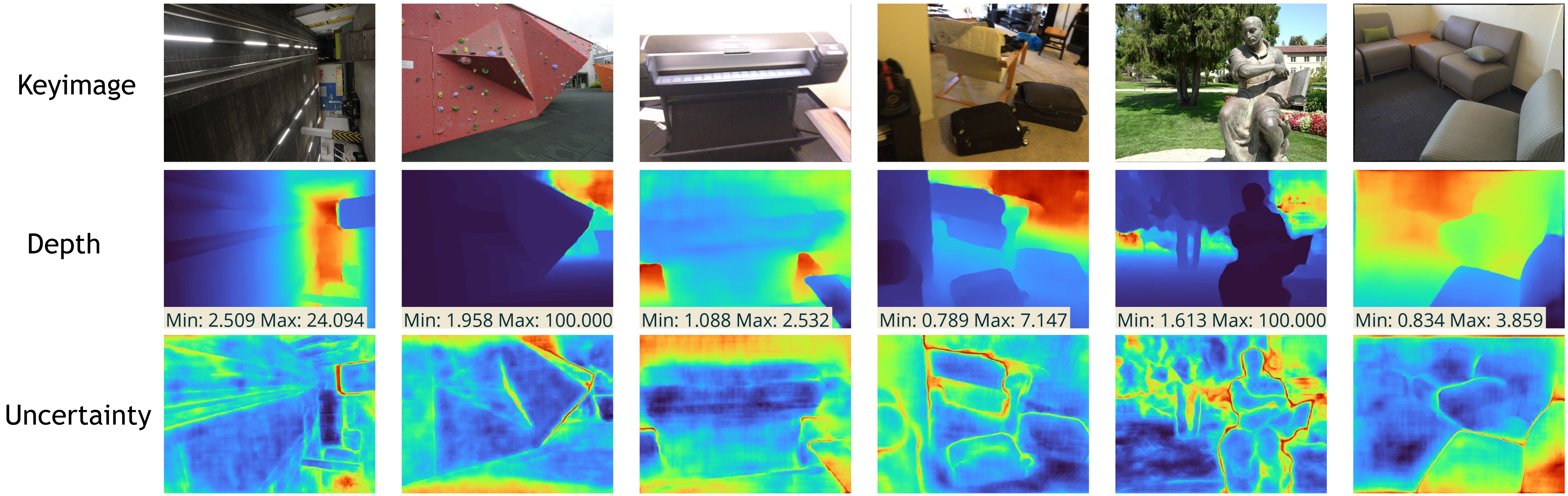
→ **Architecture:** based on DispNet

→ **Training data:** BlendedMVS + StaticThings3D

→ **Data augmentation:** Scale augmentation



Robust MVD Baseline Model
Qualitative Results



robustmvd Framework

https://github.com/lmb-freiburg/robustmvd

→ **data:** provides scripts and dataloaders to setup and use multiple datasets with a common interface

→ **model zoo:** of pre-trained depth estimation models with a common interface

→ **evaluation:** provides code to evaluate models on the rmvd benchmark

Code

robustmvd Framework

https://github.com/lmb-freiburg/robustmvd

→ **inference**: provides scripts to run multi-view depth estimation on custom data with any of the available models

→ **data viewer**: provides a data viewer to visualize input data and model predictions

→ current student project: add training code and conduct analysis

Code

Summary

→ We **show problems of current multi-view depth models:** cross-domain generalization, uncertainty estimation, multi-view fusion

→ We introduce a **benchmark to improve upon these problems**

→ **Robust MVD Baseline model** can be used as baseline on the benchmark and for robust multi-view depth estimation in applications where camera poses are known

→ **robustmvd Framework** unifies datasets and models with a common interface

Thanks!

Questions?

Code:
https://github.com/lmb-freiburg/robustmvd

KIDELTA LEARNING
Scalable AI for Automated Driving

Philipp Schröppel | Uni Freiburg

schroepp@cs.uni-freiburg.de

KI Delta Learning is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Climate Action.

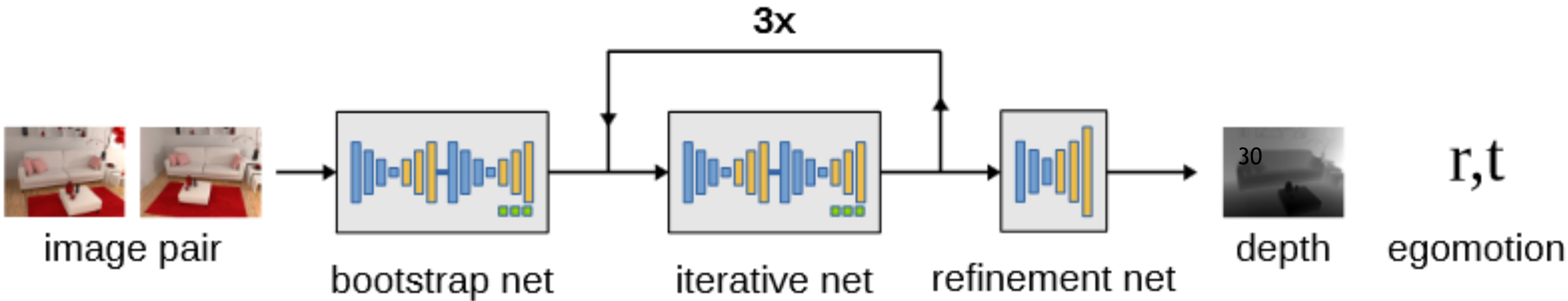
KI FAMILIE

Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag

www.ki-deltalearning.de @KI_Familie KI Familie

Related Work
DeMoN (first depth-from-”video”)

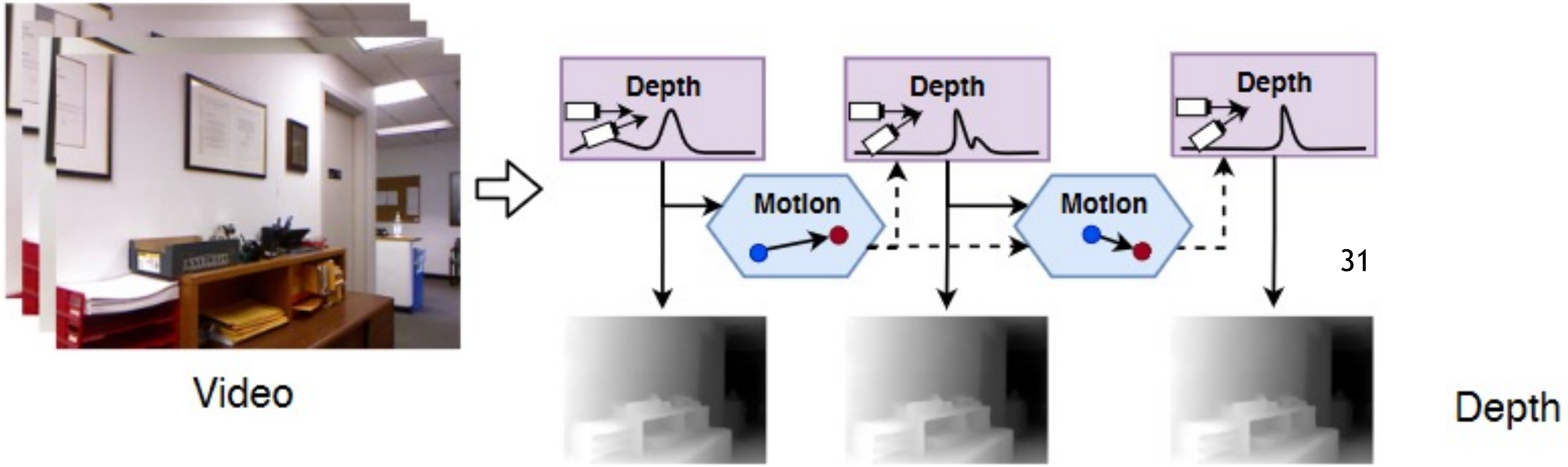
- estimates depths and poses from two views
- no explicit correlation layer within the network, but intermediate optical flow estimation
- trained on diverse data



[Ummenhofer et al. *Demon: Depth and Motion Network for Learning Monocular Stereo*. CVPR 2017]

Related Work
DeepV2D (depth-from-video)

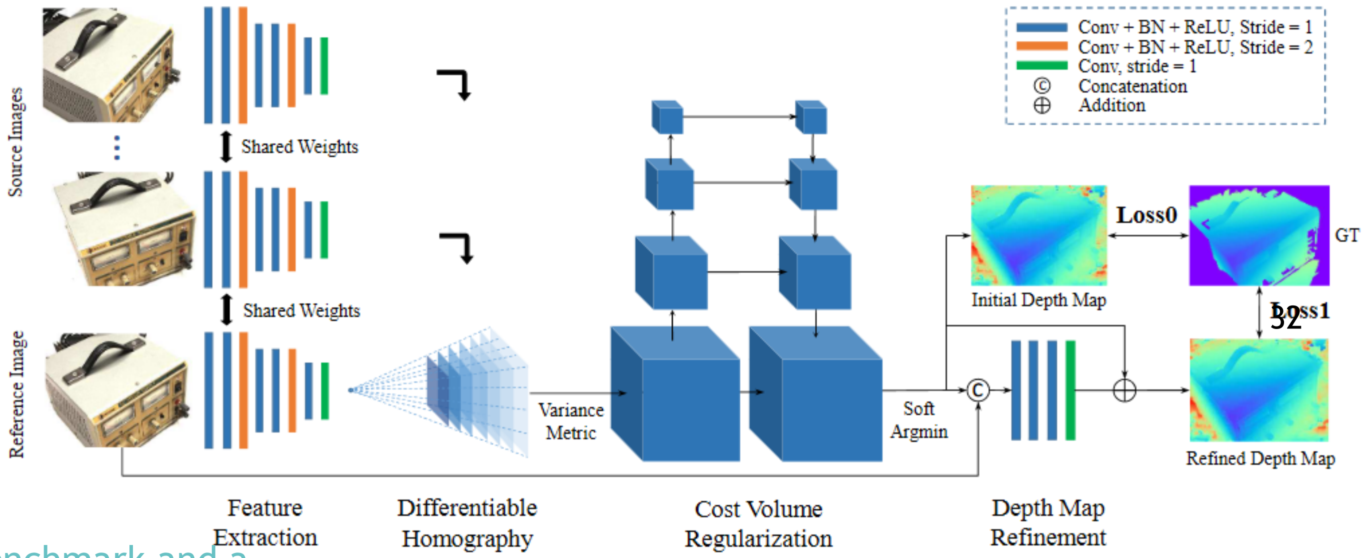
- separate mapping and tracking modules to estimate depths and poses alternatingly
- builds plane sweep stereo costvolume from learned features
- trained



[Teed and Deng.: *DeepV2D: Video to Depth with Differentiable Structure from Motion*, ICLR 2020]

Related Work
MVSNet (multi-view stereo)

- builds costvolume in a plane sweep stereo fashion based on the variance between multi-view features + decodes costvolume with 3d convolutions
- trained on B



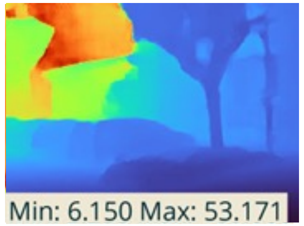
[Yao et al.: *MVSNet: Depth Inference for Unstructured Multi-view Stereo*, ECCV 2018]

Robust Multi-view Depth Estimation

Input:

Keyview

Output:



Depth map
for the keyview

Robust Multi-view Depth Estimation

We define a benchmark based on **diverse existing datasets:**

KITTI	**ScanNet**	**ETH3D**	**DTU**	**T & T**
				
Driving	Indoor	In- & outdoor	Tabletop	In- & outdoor

→ evaluation in a zero-shot cross-dataset fashion

Robust Multi-view Depth Estimation

→ We **show problems of current multi-view depth models**: cross-domain generalization, uncertainty estimation, multi-view fusion

→ The benchmark can be used **to improve upon these problems**

→ Introduce **model** for robust multi-view depth estimation on data from different domains