ECCV'20 ONLINE

23-28 AUGUST 2020

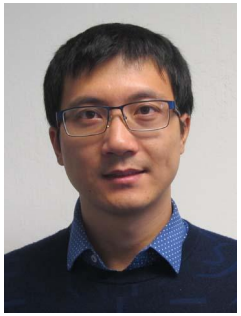
16TH EUROPEAN CONFERENCE ON

COMPUTER VISION

WWW.ECCV2020.EU

ECCV'20 ONLINE
23-28 AUGUST 2020

Pseudo RGB-D for Self-Improving Monocular SLAM and Depth Prediction



Lokender Tiwari | Pan Ji | Quoc-Huy Tran | Bingbing Zhuang | Saket Anand | Manmohan Chandraker

Presenter: Lokender Tiwari, Ph.D. Candidate at IIIT-Delhi

Project Page: https://lokender.github.io/self-improving-SLAM.html

INDRAPRASTHA INSTITUTE of INFORMATION TECHNOLOGY DELHI

NEC LABORATORIES AMERICA, INC.
Relentless passion for innovation

UC San Diego

Outline

- Motivation
- Demo 1
- Demo 2
- Demo 3

ECCV'20
ONLINE
23-28 AUGUST 2020

Motivation

ECCV'20
ONLINE
23-28 AUGUST 2020

Motivation

ECCV'20 ONLINE
23-28 AUGUST 2020

Motivation

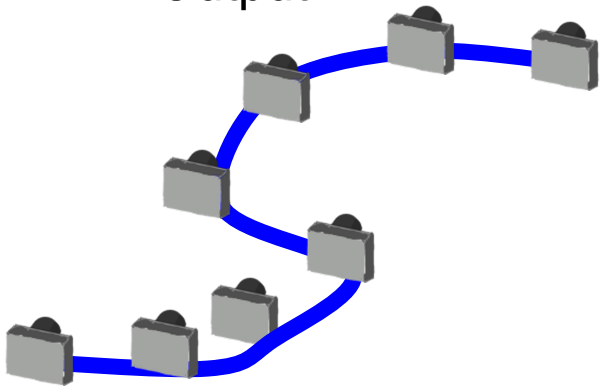
ECCV'20 ONLINE
23-28 AUGUST 2020

Motivation

RGB Images

Motivation

Output

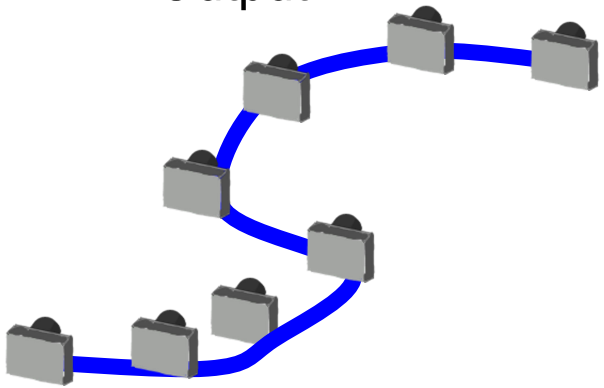


Visual Odometry
(Camera Poses)

RGB Images

Motivation

Output



Visual Odometry
(Camera Poses)
+
3D Structure
(Point Cloud)

RGB Images

Simultaneous Localization and Mapping (SLAM)

Motivation

RGB Images

Geometric RGB SLAM e.g ORB-SLAM2[1]

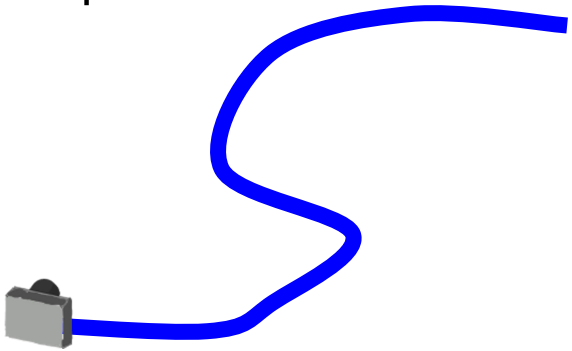
[1] Mur-Artal wt al."ORBSLAM2: An open-source slam system for monocular, stereo, and rgb-d cameras." *IEEE Transactions on Robotics* 2017

Motivation

RGB Images

Geometric RGB SLAM e.g ORB-SLAM2[1]

Output



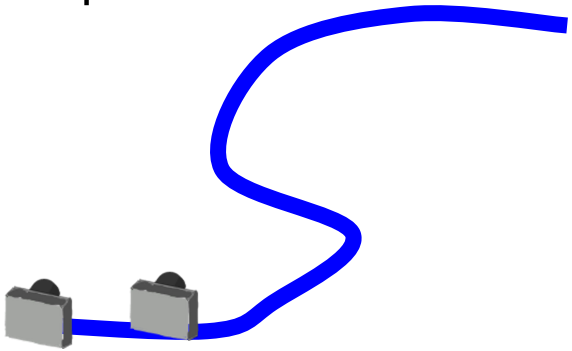
[1] Mur-Artal wt al."ORBSLAM2: An open-source slam system for monocular, stereo, and rgb-d cameras." *IEEE Transactions on Robotics* 2017

Motivation

RGB Images

Geometric RGB SLAM e.g ORB-SLAM2[1]

Output



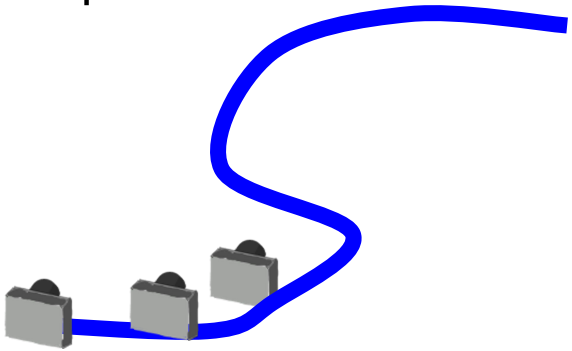
[1] Mur-Artal wt al."ORBSLAM2: An open-source slam system for monocular, stereo, and rgb-d cameras." *IEEE Transactions on Robotics* 2017

Motivation

RGB Images

Geometric RGB SLAM e.g ORB-SLAM2[1]

Output



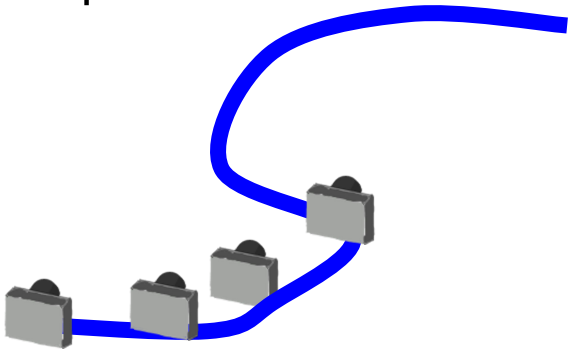
[1] Mur-Artal wt al."ORBSLAM2: An open-source slam system for monocular, stereo, and rgb-d cameras." *IEEE Transactions on Robotics* 2017

Motivation

RGB Images

Geometric RGB SLAM e.g ORB-SLAM2[1]

Output



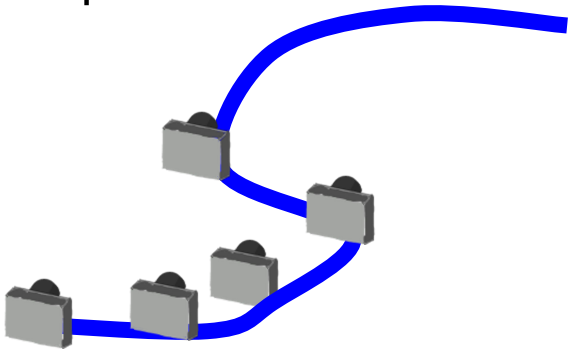
[1] Mur-Artal wt al."ORBSLAM2: An open-source slam system for monocular, stereo, and rgb-d cameras." *IEEE Transactions on Robotics* 2017

Motivation

RGB Images

Geometric RGB SLAM e.g ORB-SLAM2[1]

Output

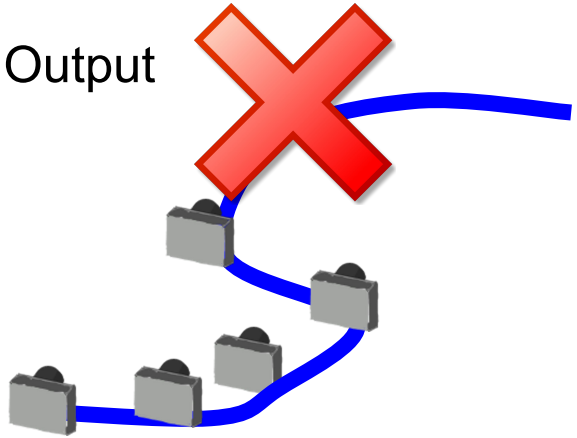


[1] Mur-Artal wt al."ORBSLAM2: An open-source slam system for monocular, stereo, and rgb-d cameras." *IEEE Transactions on Robotics* 2017

Motivation

RGB Images

Geometric RGB SLAM e.g ORB-SLAM2[1]

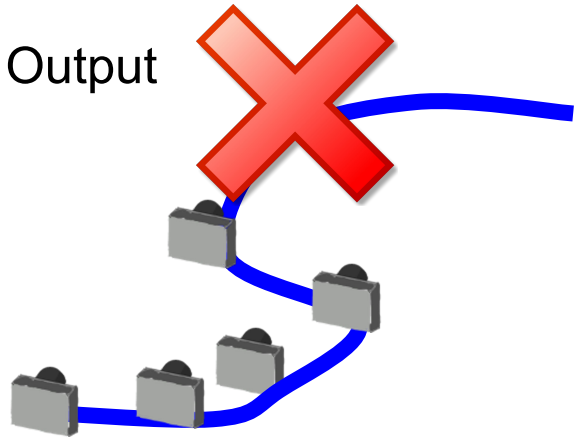


[1] Mur-Artal wt al."ORBSLAM2: An open-source slam system for monocular, stereo, and rgb-d cameras." *IEEE Transactions on Robotics* 2017

Motivation

RGB Images

Geometric RGB SLAM e.g ORB-SLAM2[1]



Suffers from:

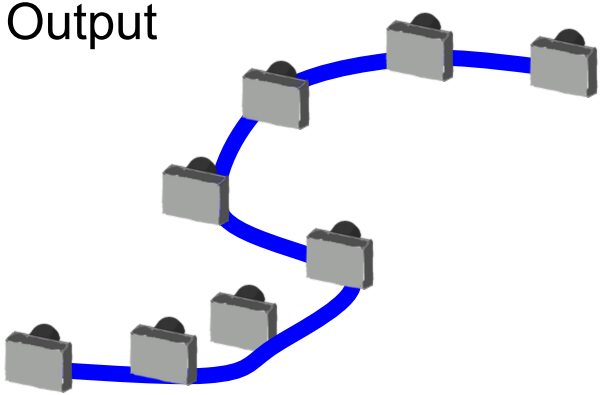
- Pure Rotational Motion
- Scale ambiguity/drift
- …

Motivation

RGB Images

RGB-D SLAM

Output

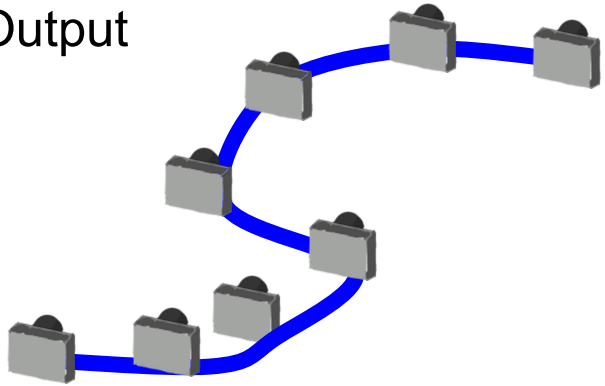


Motivation

RGB Images

RGB-D SLAM

Output



- Robust and Accurate compared to RGB-SLAM

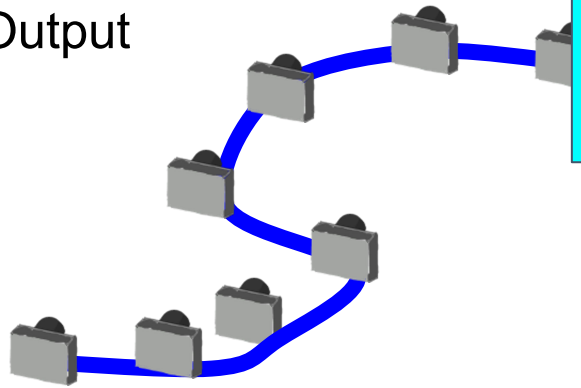
Motivation

RGB Images

RGB-D SLAM

Depth (D) is from Active depth sensor (e.g LiDAR)

Output



- Robust and Accurate compared to RGB-SLAM

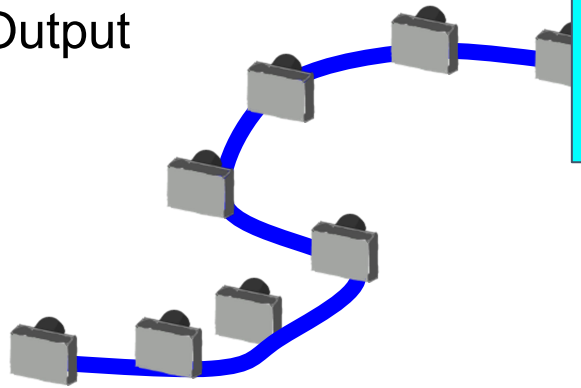
Motivation

RGB Images

RGB-D SLAM

Depth (D) is from Active depth sensor (e.g LiDAR)

Output



- Robust and Accurate compared to RGB-SLAM

Motivation

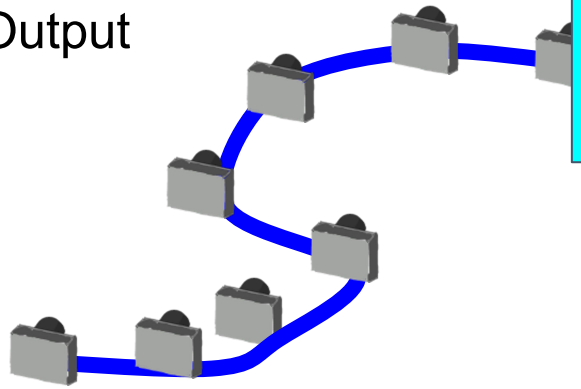
RGB Images

Pseudo RGB-D SLAM

RGB-D SLAM

Depth (D) is from Active depth sensor (e.g LiDAR)

Output

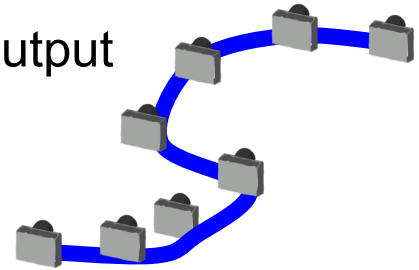


- Robust and Accurate compared to RGB-SLAM

Motivation

RGB Images

Output



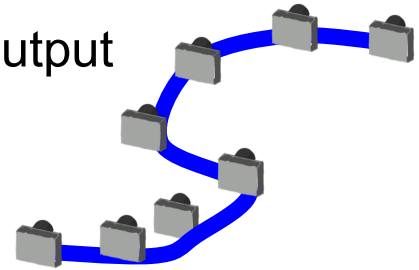
Pseudo RGB-D SLAM

- Use unsupervised CNN based depth prediction model as a Pseudo Active depth sensor.

Motivation

RGB Images

Output



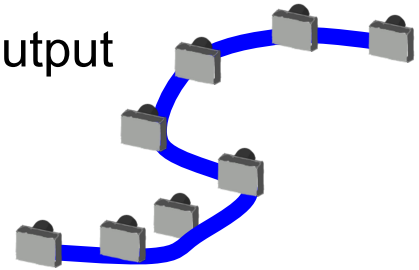
Pseudo RGB-D SLAM

- Use unsupervised CNN based depth prediction model as a Pseudo Active depth sensor.
- Is it straightforward ?

Motivation

RGB Images

Output



Pseudo RGB-D SLAM

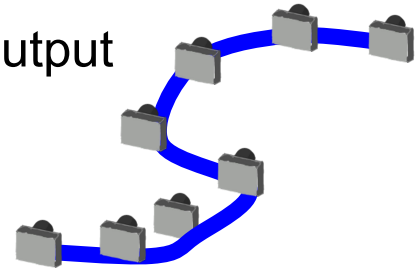
- Use unsupervised CNN based depth prediction model as a Pseudo Active depth sensor.
- Is it straightforward ?

NO

Motivation

RGB Images

Output



Pseudo RGB-D SLAM

- Use unsupervised CNN based depth prediction model as a Pseudo Active depth sensor.
- Is it straightforward ?

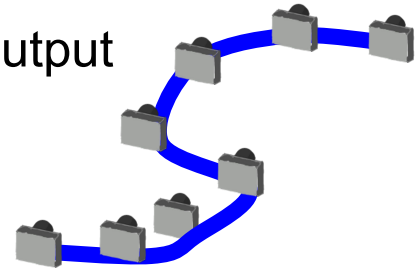
NO

- Because CNN predicts depth maps at very different metric scales.

Motivation

RGB Images

Output



Pseudo RGB-D SLAM

- Use unsupervised CNN based depth prediction model as a Pseudo Active depth sensor.
- Is it straightforward ?

NO

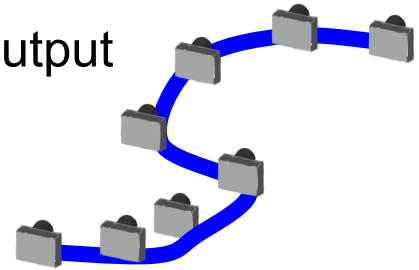
- Because CNN predicts depth maps at very different metric scales.

Adaptations, e.g., Adaptive baseline

Motivation

RGB Images

Output



Pseudo RGB-D SLAM

- Pseudo RGB-D SLAM performance depends on quality of Depth Maps.

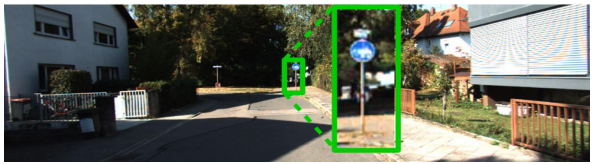
Motivation

RGB Images

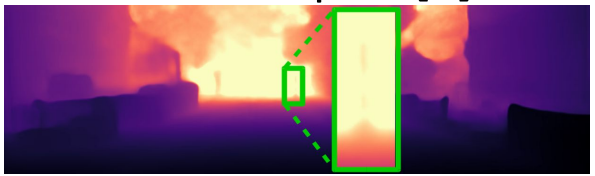
Pseudo RGB-D SLAM

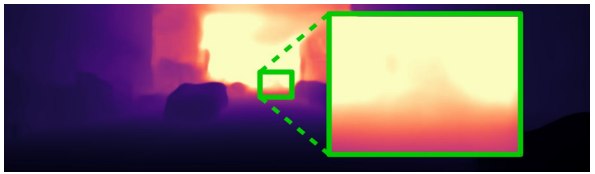
- Pseudo RGB-D SLAM performance depends on quality of Depth Maps.

RGB

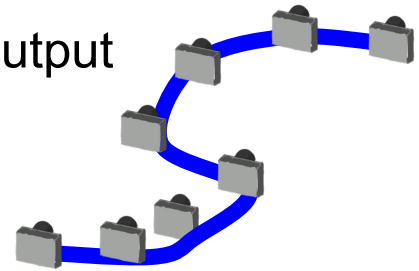


MonoDepth2[1]





Output

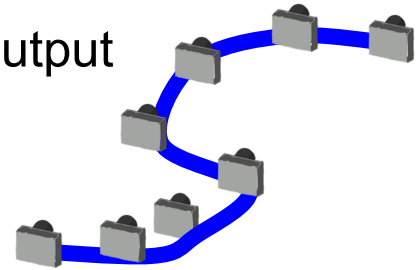


[1] Godard, Clément, et al. "Digging into self-supervised monocular depth estimation." *in ICCV 2019*

Motivation

RGB Images

Output



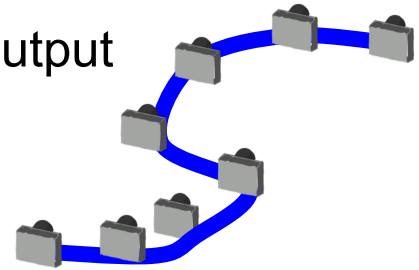
Unsupervised CNN Based Depth Prediction

- Formulate as a novel view synthesis problem
- **Depth** + **Pose** Network
- Quality of depth estimates depends on quality of poses from pose estimation network
- Pose network often perform poorly

Motivation

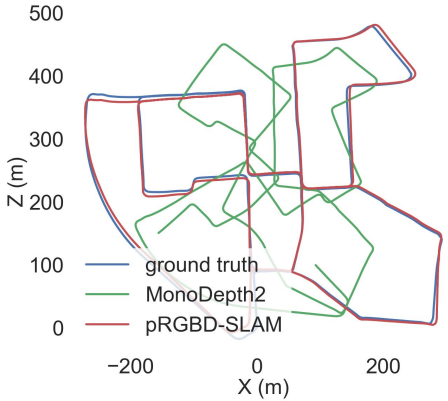
RGB Images

Output



Unsupervised CNN Based Depth Prediction

- Formulate as a novel view synthesis problem.
- **Depth** + **Pose** Network
- Quality of depth estimates depends on quality of poses from pose estimation network.
- Pose network often perform poorly.



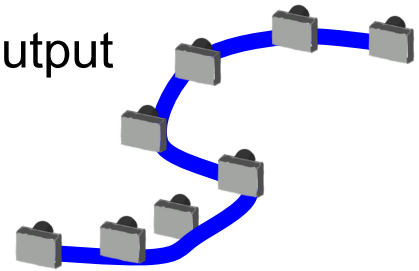
Motivation

RGB Images

Unsupervised CNN Based Depth Prediction

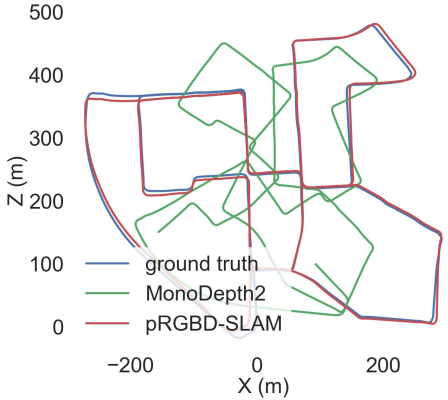
- Formulate as a novel view synthesis problem.
- **Depth** + **Pose** Network
- Quality of depth estimates depends on quality of poses from pose estimation network.
- Pose network often perform poorly.

Output



Does not model:

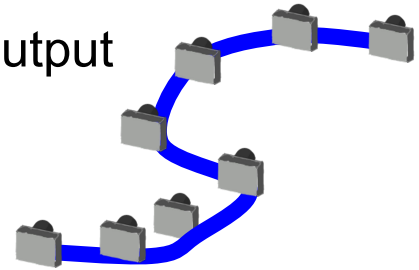
- Photo changes
- Wide-baseline constraints (beyond 3-5 frames)
- ....



Motivation

RGB Images

Output



Pseudo RGB-D SLAM

- Quality of **camera poses** depends on the quality of **depth maps** from CNN.

Monocular Depth Prediction

- Quality of **depth maps** depends on quality of **camera poses** from pose network

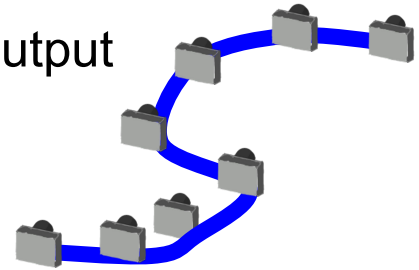
geometric-CNN Framework

We propose a Self-Supervised, Self-Improving framework.

Motivation

RGB Images

Output



Pseudo RGB-D SLAM

- Quality of **camera poses** depends on the quality of **depth maps** from CNN.

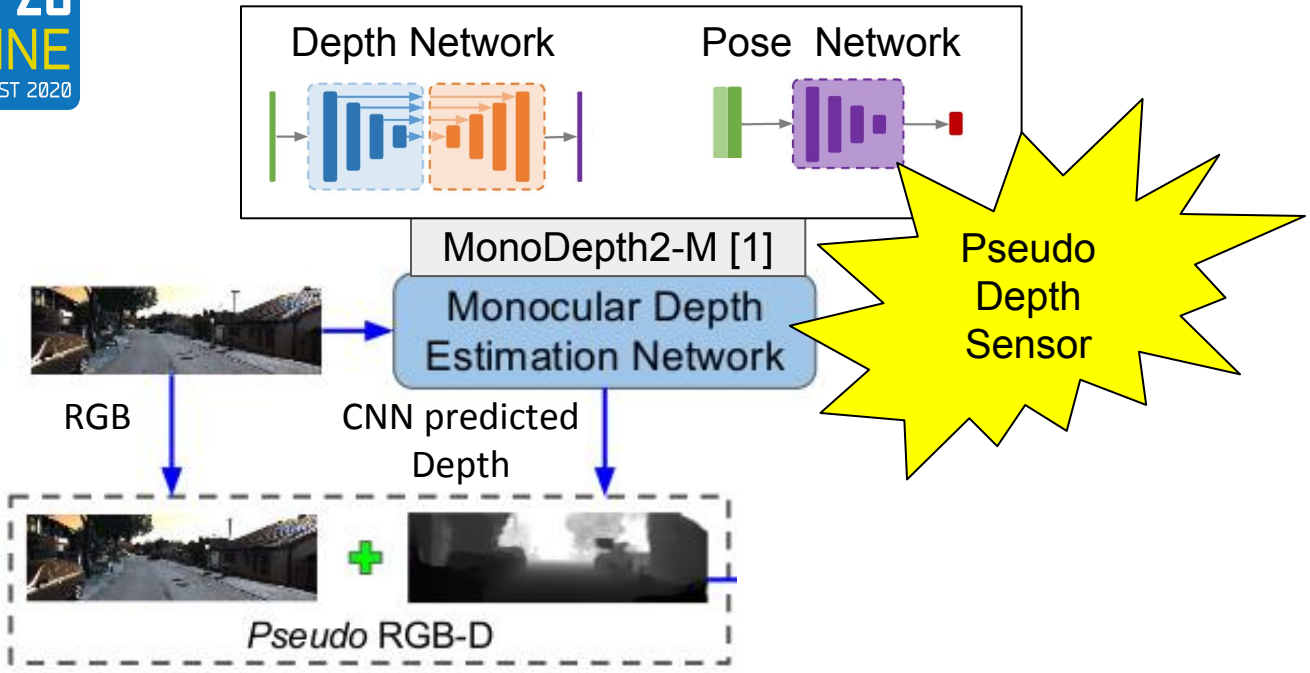
Monocular Depth Prediction

- Quality of **depth maps** depends on quality of **camera poses** from pose network

geometric-CNN Framework

We propose a Self-Supervised, Self-Improving framework.

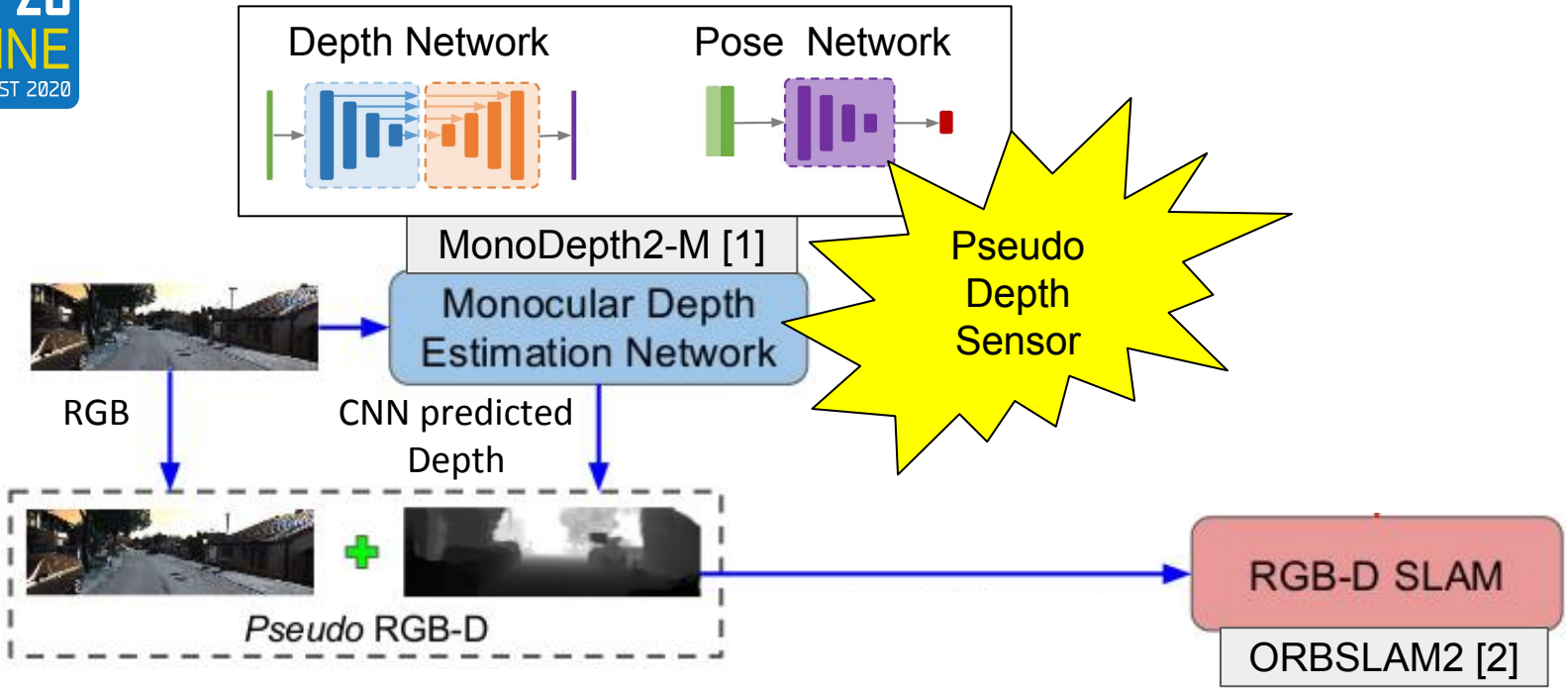
A Self-Supervised, Self-Improving Framework



- Prepare **Pseudo RGB-D** data

[1] Godard, Clément, et al. "Digging into self-supervised monocular depth estimation." *in ICCV 2019*

A Self-Supervised, Self-Improving Framework

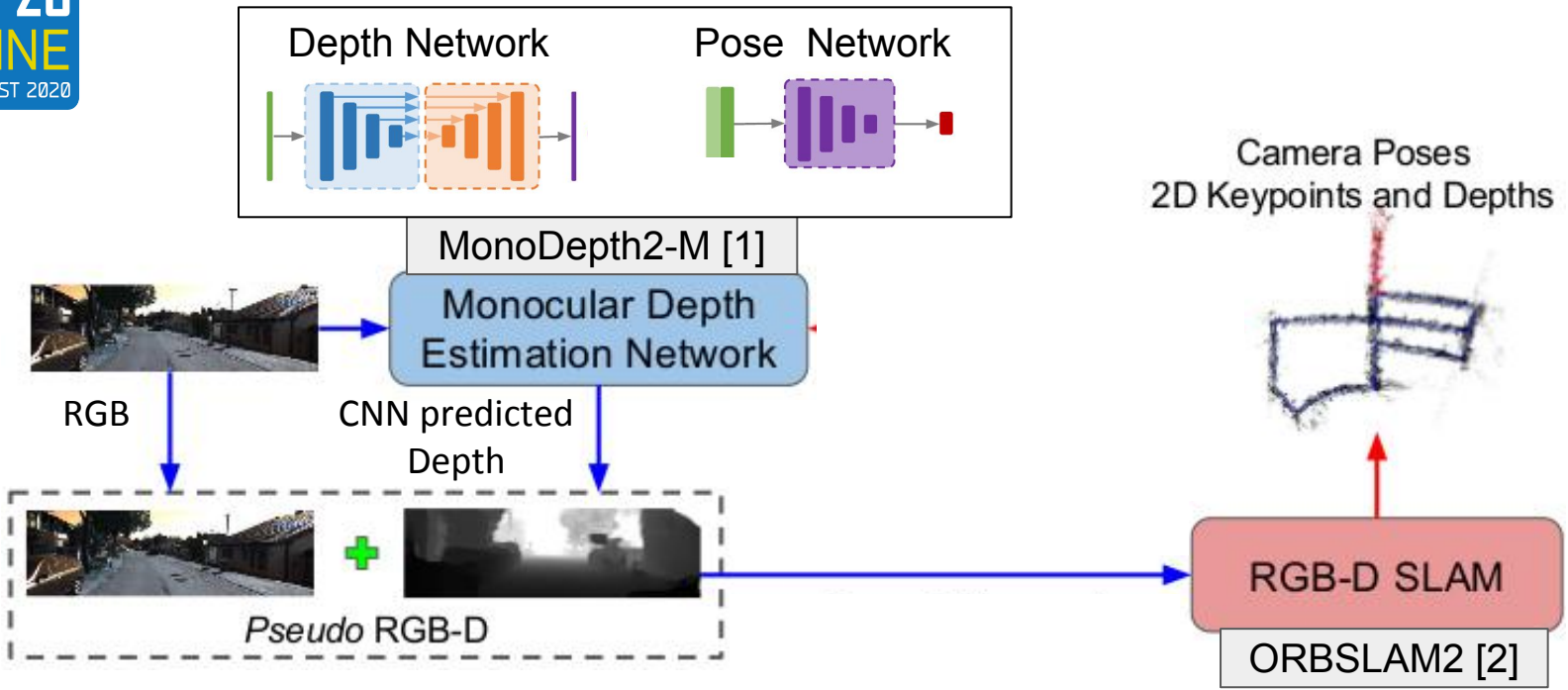


- Prepare **Pseudo RGB-D** data
- Run RGB-D SLAM on **Pseudo RGB-D** pairs. We use RGB-D version of **ORB-SLAM2 [2]** as base RGB-D SLAM

[1] Godard, Clément, et al. "Digging into self-supervised monocular depth estimation." *in ICCV 2019*

[2] Mur-Artal wt al."ORBSLAM2: An open-source slam system for monocular, stereo, and rgb-d cameras." *IEEE Transactions on Robotics* 2017

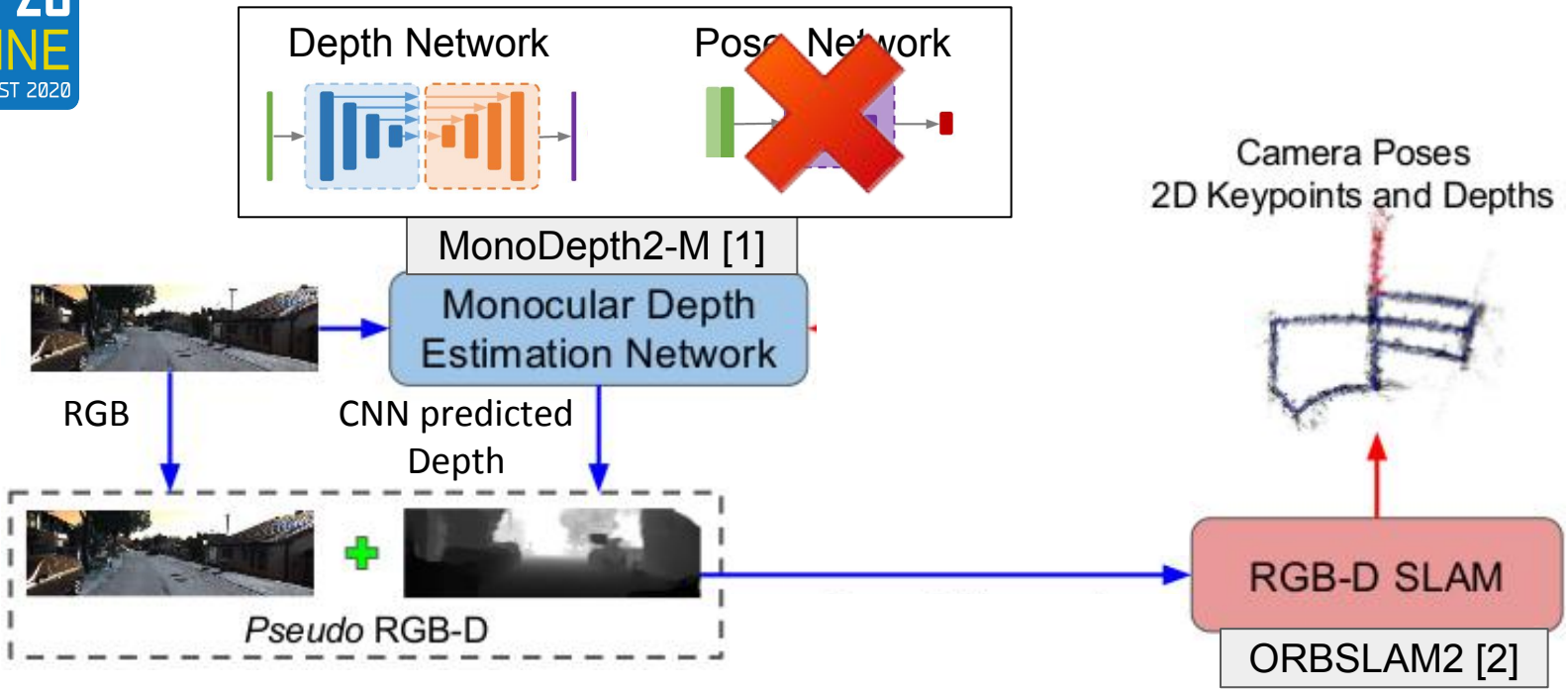
A Self-Supervised, Self-Improving Framework



- Prepare **Pseudo RGB-D** data
- Run RGB-D SLAM on **Pseudo RGB-D** pairs. We use RGB-D version of **ORB-SLAM2 [2]** as base RGB-D SLAM
- Save Pseudo RGB-D SLAM outputs (Camera poses, keyframes, tracked keypoints and their depth values)

[1] Godard, Clément, et al. "Digging into self-supervised monocular depth estimation." *in ICCV 2019*
[2] Mur-Artal wt al."ORBSLAM2: An open-source slam system for monocular, stereo, and rgb-d cameras." *IEEE Transactions on Robotics* 2017

A Self-Supervised, Self-Improving Framework

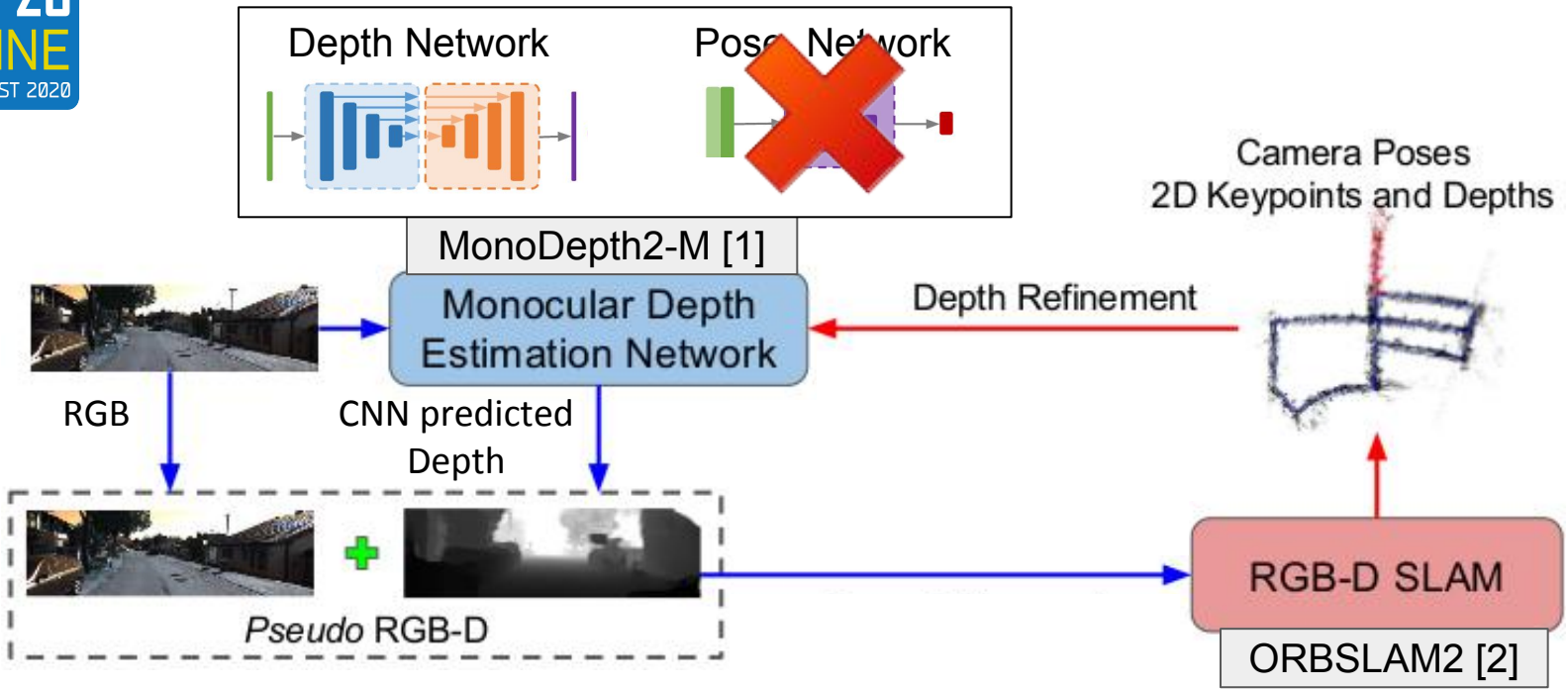


- **Depth Refinement**
  - Disable MonoDepth2 pose network
  - Use camera poses obtained from Pseudo RGB-D SLAM

[1] Godard, Clément, et al. "Digging into self-supervised monocular depth estimation." *in ICCV 2019*

[2] Mur-Artal wt al."ORBSLAM2: An open-source slam system for monocular, stereo, and rgb-d cameras." *IEEE Transactions on Robotics* 2017

A Self-Supervised, Self-Improving Framework

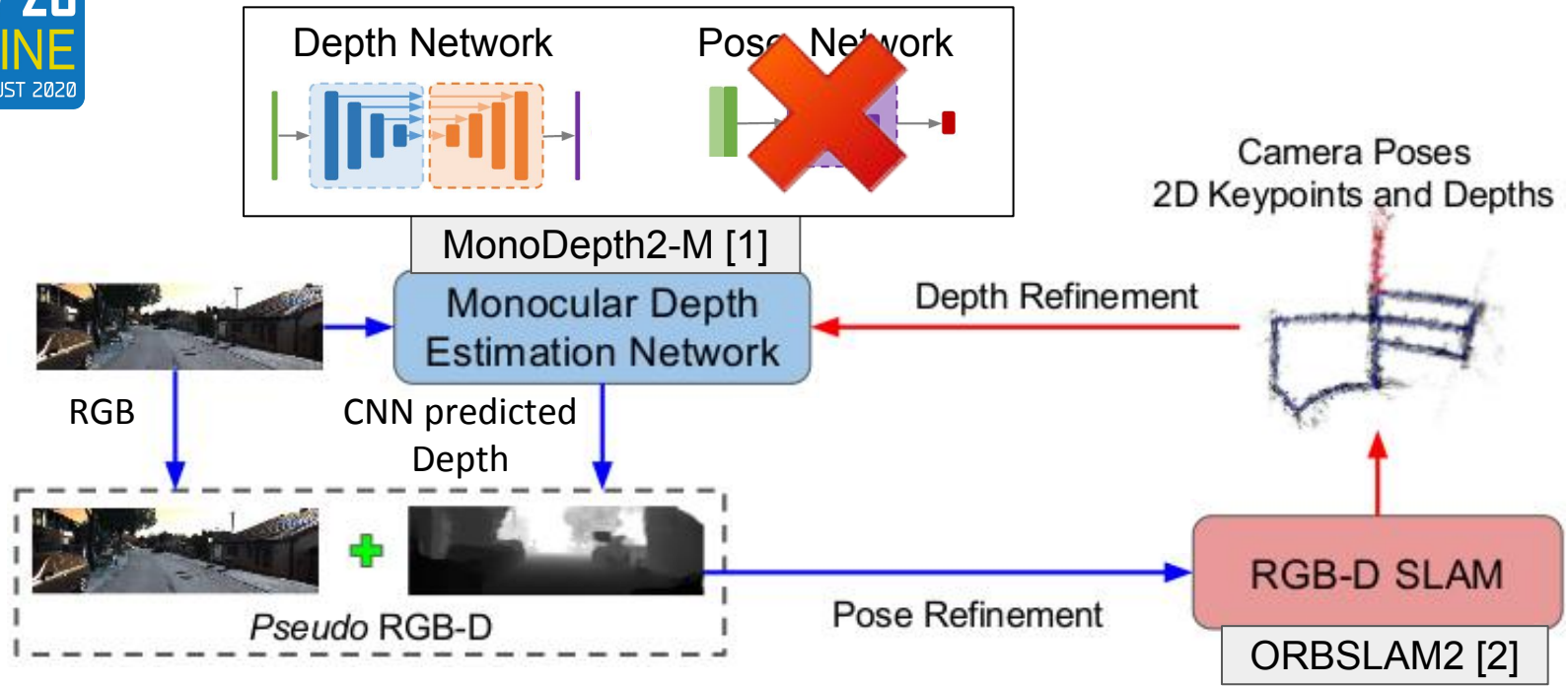


- **Depth Refinement**
  - Disable MonoDepth2 pose network
  - Use camera poses obtained from Pseudo RGB-D SLAM

[1] Godard, Clément, et al. "Digging into self-supervised monocular depth estimation." *in ICCV 2019*

[2] Mur-Artal wt al."ORBSLAM2: An open-source slam system for monocular, stereo, and rgb-d cameras." *IEEE Transactions on Robotics* 2017

A Self-Supervised, Self-Improving Framework

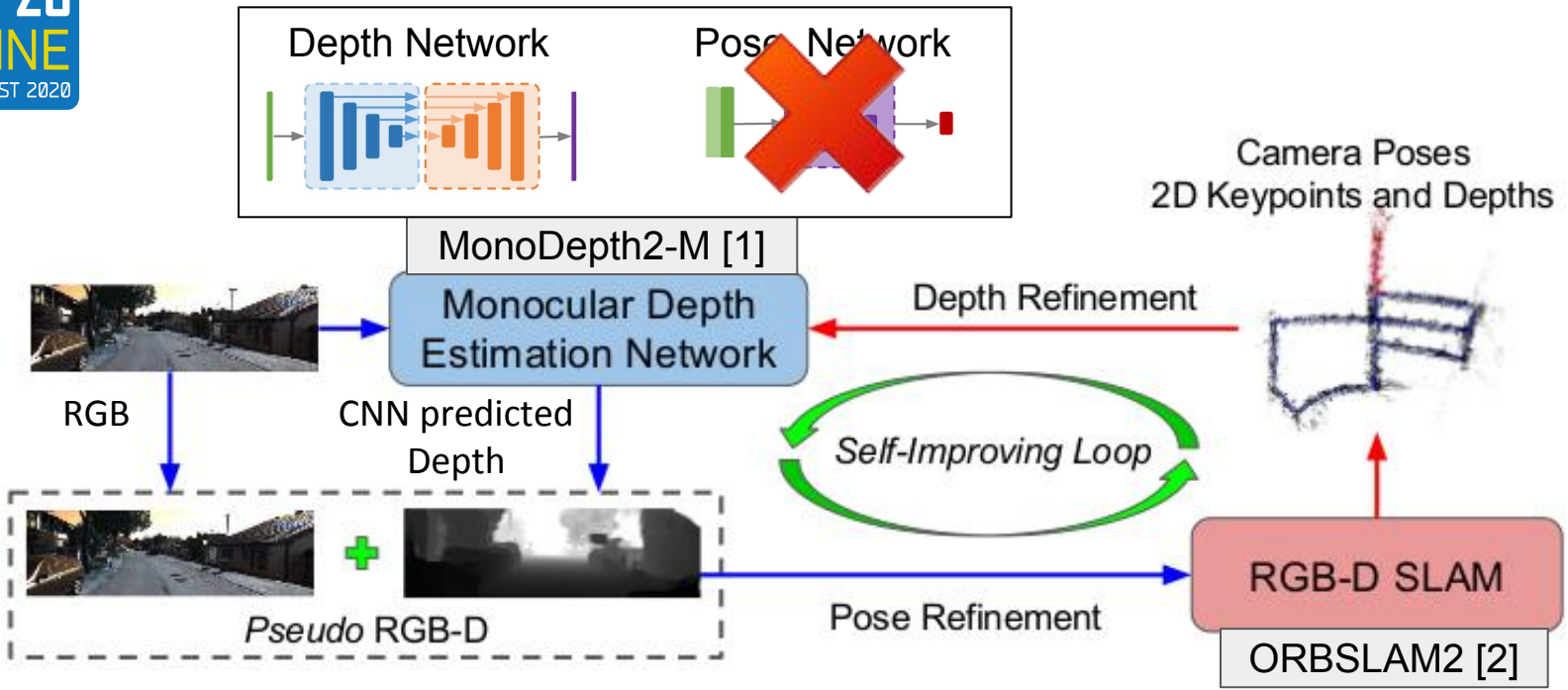


- **Pose Refinement**
  - Use the refined depth model to prepare Pseudo RGB-D data
  - Re-run Pseudo RGBD-D SLAM and get refined camera poses, keypoints amd their updated locations

[1] Godard, Clément, et al. "Digging into self-supervised monocular depth estimation." *in ICCV 2019*

[2] Mur-Artal wt al."ORBSLAM2: An open-source slam system for monocular, stereo, and rgb-d cameras." *IEEE Transactions on Robotics* 2017

A Self-Supervised, Self-Improving Framework



- **Self-Improving Loop**

[1] Godard, Clément, et al. "Digging into self-supervised monocular depth estimation." *in ICCV 2019*
[2] Mur-Artal wt al."ORBSLAM2: An open-source slam system for monocular, stereo, and rgb-d cameras." *IEEE Transactions on Robotics* 2017

DEMO 1

KITTI Odometry Sequence 01
(1100 Frames)

DEMO 1

KITTI Odometry Sequence 01 (1100 Frames)

Sequence	Dimension (m×m)	ORB-SLAM KFs	ORB-SLAM RMSE (m)	+ Global BA (20 its.) RMSE (m)	+ Global BA (20 its.) Time BA (s)
KITTI 00	564 × 496	1391	6.68	5.33	24.83
KITTI 01	1157 × 1827	X	X	X	X
KITTI 02	599 × 946	1801	21.75	21.28	30.07

[1] Mur-Artal wt al."ORBSLAM2: An open-source slam system for monocular, stereo, and rgb-d cameras." *IEEE Transactions on Robotics* 2017

DEMO 2

KITTI Odometry Sequence 19
(4985 Frames)

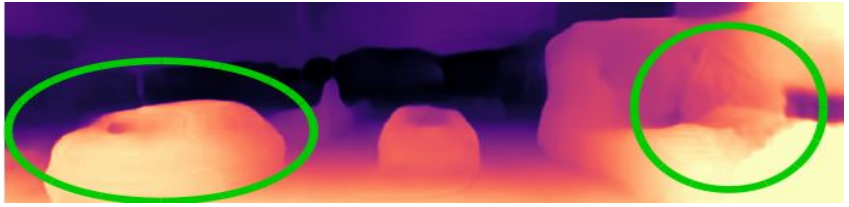
DEMO 3

KITTI Eigen Split Test Set
(Improved Depth Estimates)
(Qualitative Results)

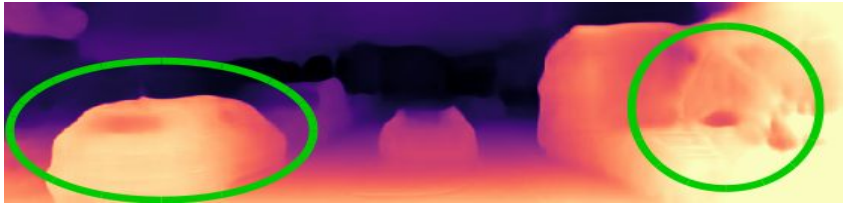
DEMO 3

RGB

MonoDepth2 [1]-Stereo Supervision



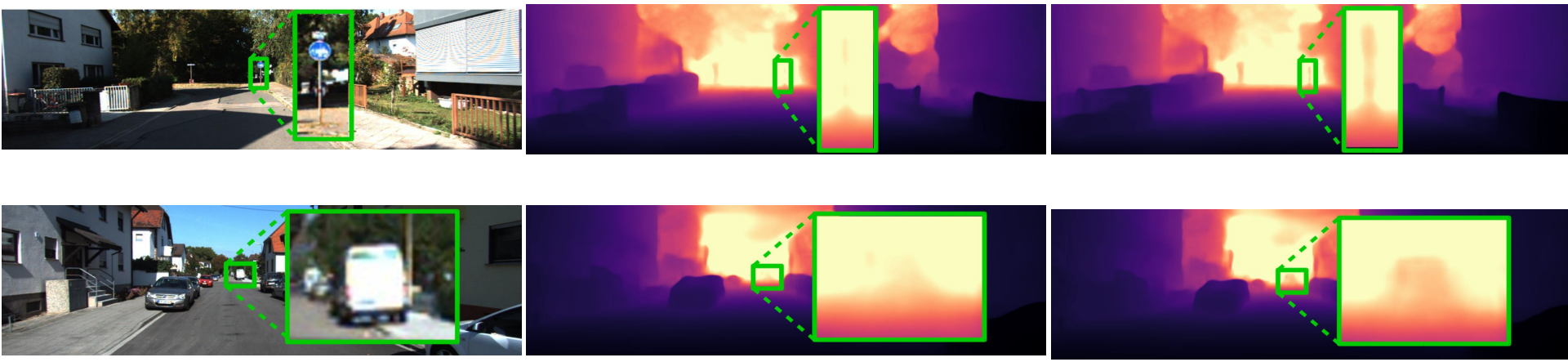
MonoDepth2 [1]-Monocular Supervision



pRGBD-Refined (Proposed Method)

[1] Godard, Clément, et al. "Digging into self-supervised monocular depth estimation." *in ICCV 2019*

DEMO 3



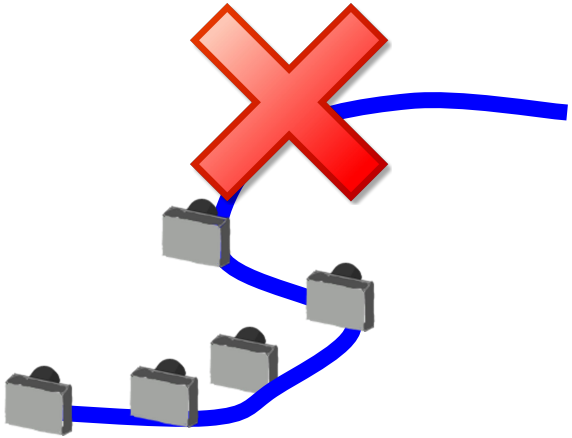
RGB | MonoDepth2[1]-Monocular Supervision | pRGBD-Refined (Proposed Method)

- Visual improvements in the depth of farther points.

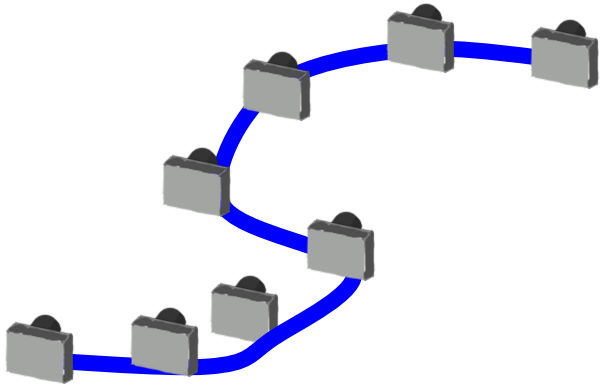
Motivation

RGB Images

Geometric RGB SLAM e.g ORB-SLAM2

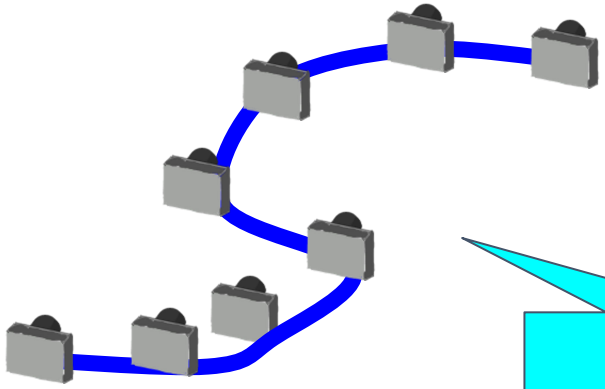


Motivation



Motivation

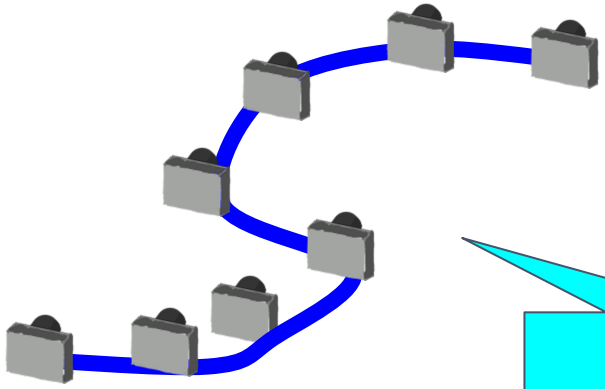
RGB-D SLAM e.g ORB-SLAM2



Depth (D) from
Active depth sensor
(e.g LiDAR)

Motivation

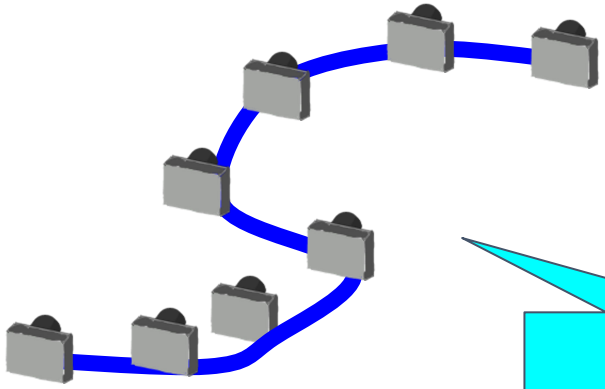
RGB-D SLAM e.g ORB-SLAM2



Depth (D) from
Active depth sensor
(e.g LiDAR)

Motivation

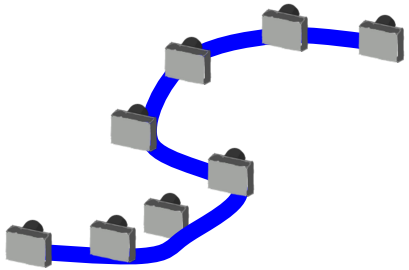
RGB-D SLAM e.g ORB-SLAM2



Depth (D) from
Active depth sensor
(e.g LiDAR)

Pseudo RGB-D SLAM

Motivation



Pseudo RGB-D SLAM

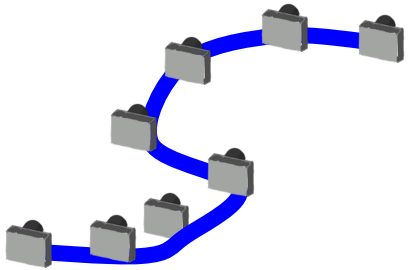
- Use CNN based depth estimation model as a Pseudo Active depth sensor.
- Is it straightforward ?

NO

- Because CNN predicts depth maps at very different metric scales.

Adaptations, e.g., Adaptive baseline

Motivation



Pseudo RGB-D SLAM

- Use CNN based depth estimation model as a Pseudo Active depth sensor.
- Is it straightforward ?

NO

- Because CNN predicts depth maps at very different metric scales.

Adaptations, e.g., Adaptive baseline