ICSCC 2021

1-3 July 2021
Kochi, India

8th International Conference on Smart Computing and Communications

IEEE

Paper ID: 314

Title: A Comparative Study of Deep Learning-based Depth Estimation Approaches: Application to Smart Mobility

Name of Authors: Antoine Mauri - Redouane Khemmar - Benoit Decoux - Tahar Benmoumen - Madjid Haddad - Rémi Boutteau

Presented By: Redouane Khemmar
Affiliation: Normandie Univ, UNIROUEN, ESIGELEC, IRSEEM

Outline

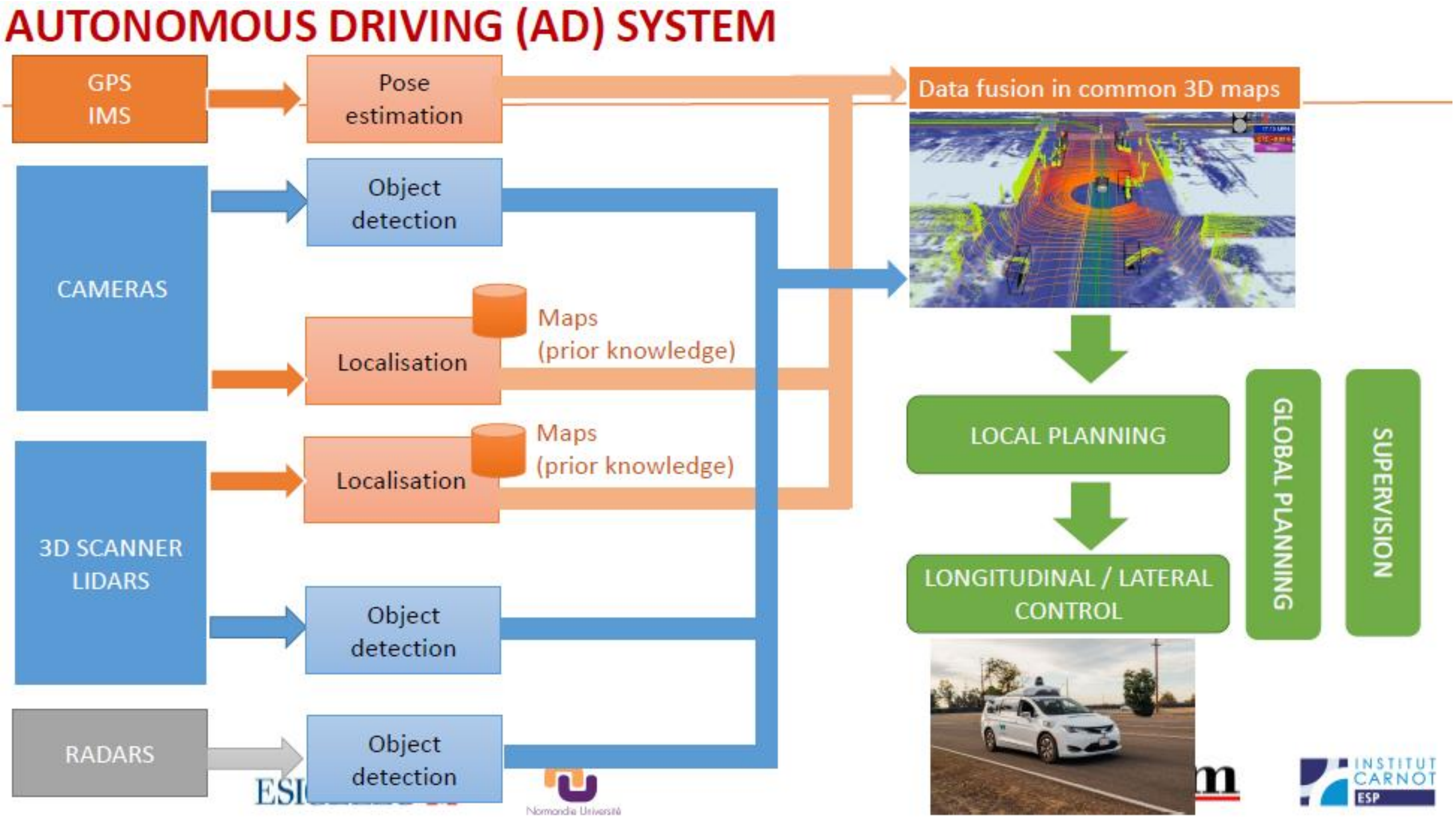
Context & Motivation

- Perception of environment is an important aspect for Autonomous Vehicles (AV)
  - Localization of potential obstacles is crucial to prevent collisions for AV
  - Object detection, localization, and tracking are important for scene analysis and understanding

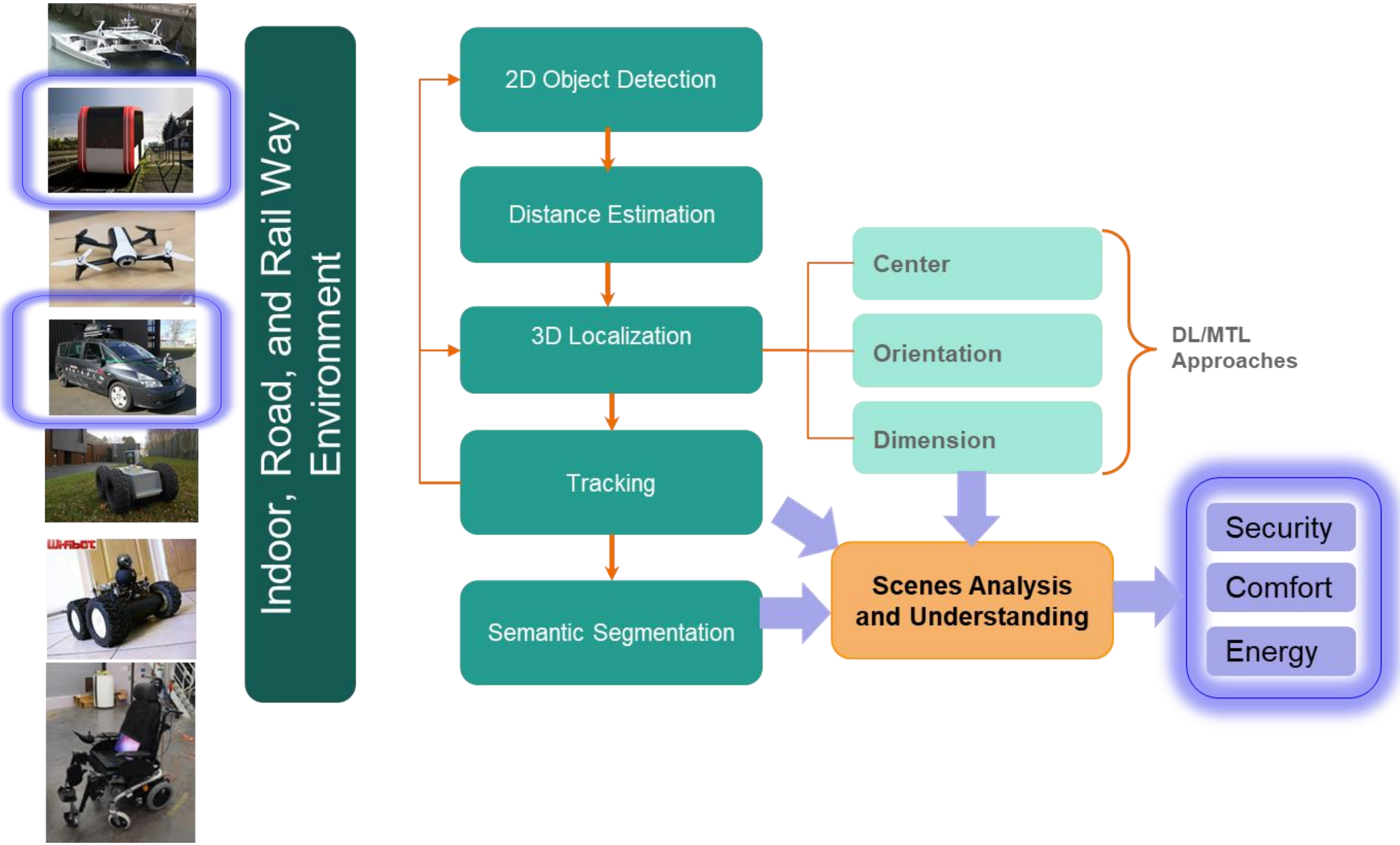
- CNN based Depth Estimation Methods allow to use Images from a Monocular or Stereoscopic Camera to get a Dense Depth Prediction
  - Expensive depth sensors like LiDAR could be replaced by a single camera

- Few works has been done to compare the performances under realistic environment

Context & Motivation



Context & Motivation



A. Mauri et al. A new Evaluation Approach for Deep Learning-based Monocular Depth Estimation Methods. ICSCC 2021. 03/07/2021.

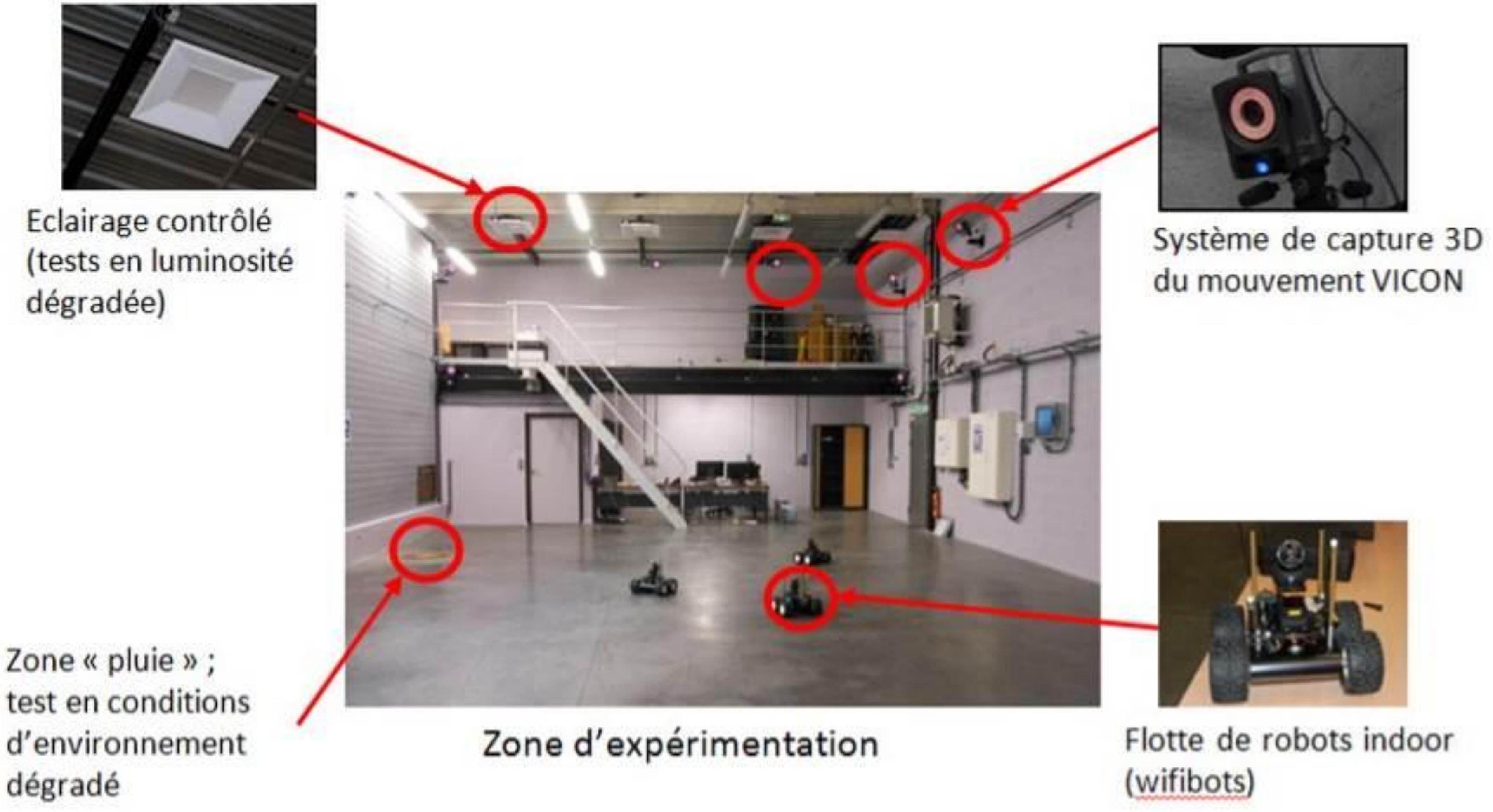
Context & Motivation

▶ **Autonomous Vehicle: Acquisition System**

A. Mauri et al. A new Evaluation Approach for Deep Learning-based Monocular Depth Estimation Methods. ICSCC 2021. 03/07/2021.

Context & Motivation

▶ Autonomous Vehicle: Autonomous Navigation Laboratory



A. Mauri et al. A new Evaluation Approach for Deep Learning-based Monocular Depth Estimation Methods. ICSCC 2021. 03/07/2021.

Related Work : Depth Evaluation Metrics

- Depth error metrics used in our study:
  - Relative Error : $RE = \frac{1}{N}\sum_i\sum_j \frac{|g_{i,j}-p_{i,j}|}{g_{i,j}}$
  - Squared Relative Error : $SRE = \frac{1}{N}\sum_i\sum_j \frac{|g_{i,j}-p_{i,j}|^2}{g_{i,j}}$
  - Root Mean Squared Error : $RMSE = \sqrt{\frac{1}{N}\sum_i\sum_j (g_{i,j}-p_{i,j})^2}$
  - Logarithmic Root Mean Squared Error : $logRMSE = \sqrt{\frac{1}{N}\sum_i\sum_j (log(g_{i,j})-log(p_{i,j}))^2}$

p: depth prediction of a pixel in the image

g: depth ground truth of a pixel in the image

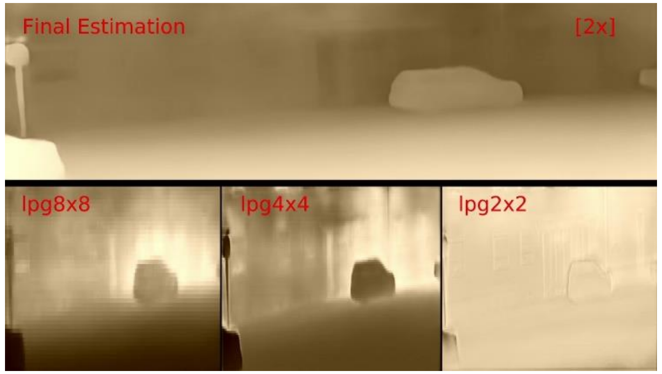
N: total of pixels in the image

A. Mauri et al. A new Evaluation Approach for Deep Learning-based Monocular Depth Estimation Methods. ICSCC 2021. 03/07/2021.

Evaluated Single Image Depth Estimation Methods

BTS[1] (Big To Small)

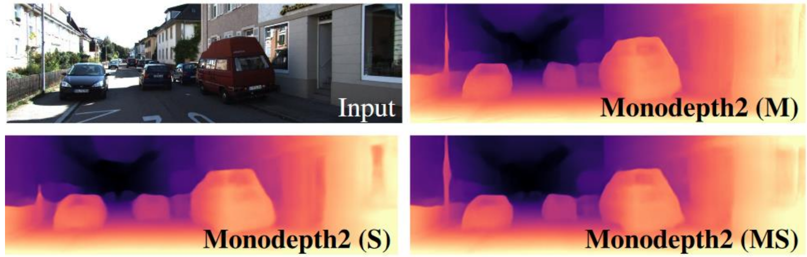
- **Supervised** monocular depth estimation
- Layers located at multiple stages of the decoding phase are used
- Their outputs are combined to **predict depth at full resolution**
- Achieves top precision on KITTI depth benchmark



BTS Approach

Monodepth2[2]

- Can be trained with or without depth supervision
- With self-supervision, the problem of depth estimation is casted into an image reconstruction one
- Self-supervision works with:
  - Monocular image sequences (M)
  - Stereo data (S)
  - Or both of them (MS)



Monodepth2 Approach

1. J. H. Lee, M.-K. Han, D. W. Ko, and I. H. Suh, "From big to small: Multi-scale local planar guidance for monocular depth estimation," arXiv preprint arXiv:1907.10326, 2019.
2. C. Godard, O. Mac Aodha, M. Firman, and G. Brostow, "Digging into self-supervised monocular depth estimation," arXiv preprint arXiv:1806.01260, 2018.

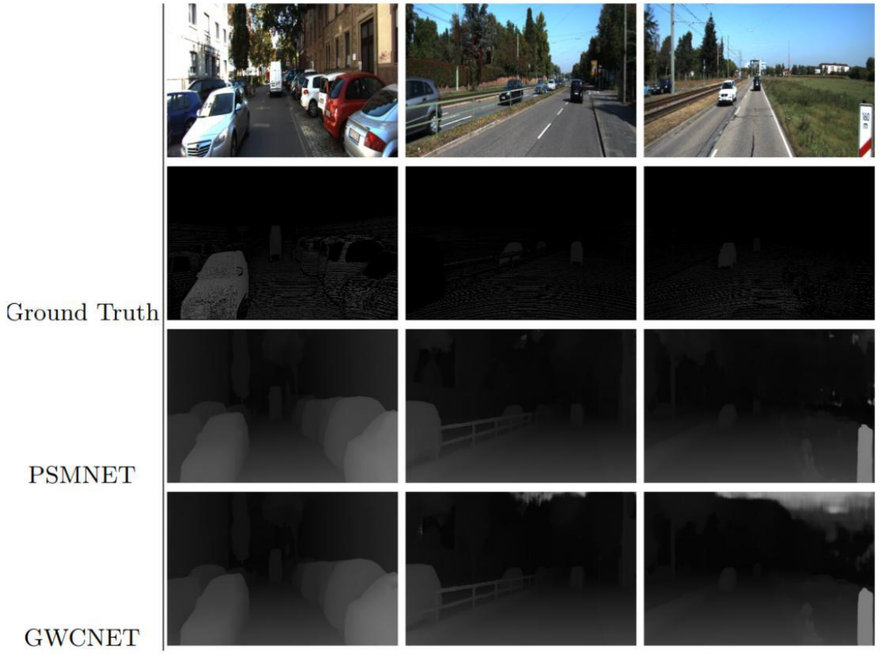
Evaluated Stereoscopic Image Depth Eestimation Methods

PSMNet[1]

- **Supervised** stereoscopic depth estimation
- Use **Spatial Pyramid Pooling** to expand the receptive field
- Global contextual informations are extracted by a stacked hourglass 3D CNN

GwcNet[2]

- **Supervised** stereoscopic depth estimation
- Group wise correlation are used for providing matching features
- Improved stacked hourglass network



PSMNET and GWCNET

1. PSMNET: Pyramid Stereo Matching Network.
2. GWCNet: Group-wise Correlation Stereo Network

1. J.-R. Chang and Y.-S. Chen, “Pyramid stereo matching network,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5410–5418, 2018.
2. X. Guo, K. Yang, W. Yang, X. Wang, and H. Li, “Group-wise correlation stereo network,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3273–3282, 2019.

Experimental Results

Dataset : KITTI[1]

- Highly used dataset for road environment with real world traffic
- Data are calibrated and synchronized and come from multiple sensors **:** Stereo camera, Velodyne Lidar, GPS/IMU navigation system
- Used as benchmark for depth estimation, optical flow, and object detection tasks

Results

	RelErr	SqRel	RMSE	logRMSE
Monodepth2	0.115	0.882	4.701	0.190
BTS	**0.060**	**0.249**	**2.798**	**0.096**

Results of single image-based methods

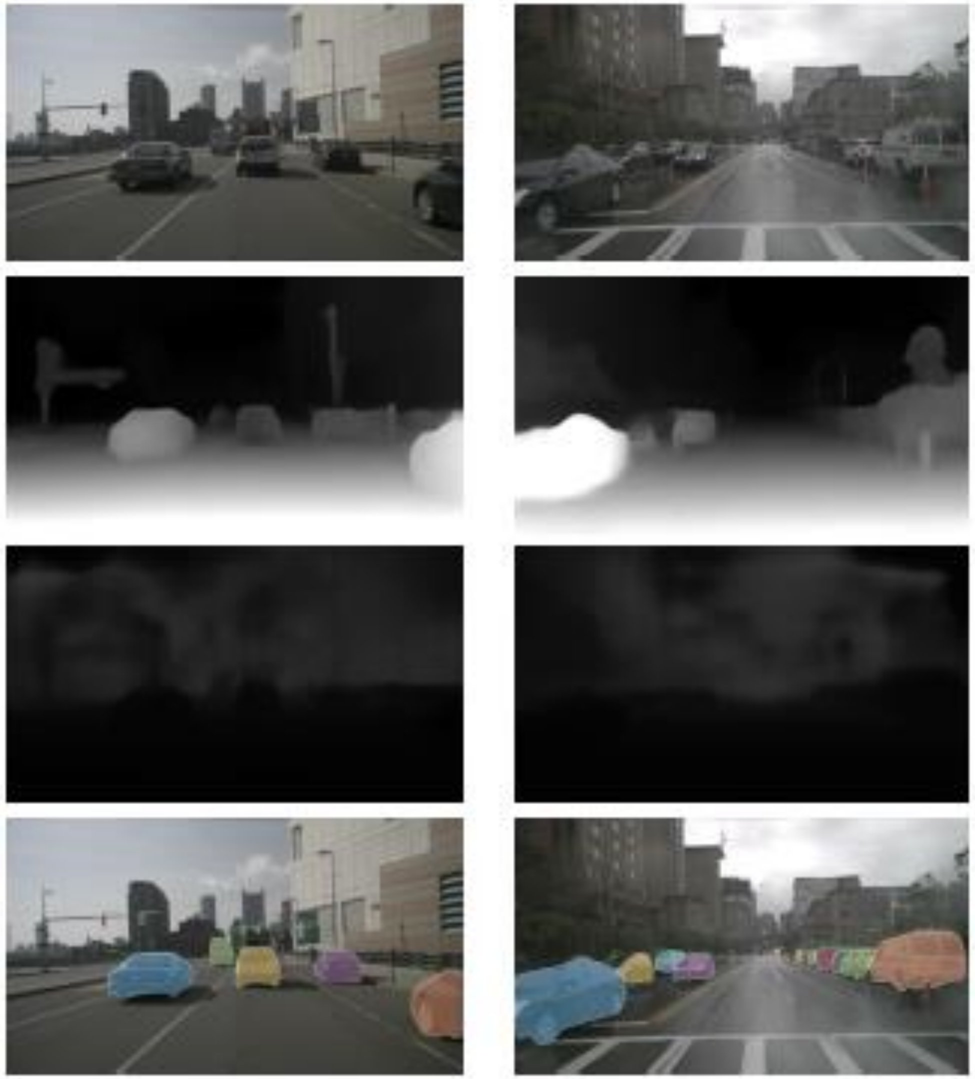
	RelErr	SqRel	RMSE	logRMSE
GWCNET	**0.018**	**0.048**	**0.981**	**0.042**
PSMNET	0.032	0.061	1.139	0.056

Results of stereoscopic-based methods

1. A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," The International Journal of Robotics Research, vol. 32, no. 11, pp. 1231–1237, 2013.

Experimental Results

- **Input data of our Object Distance Evaluation Protocol:**
  - Input image fed to the depth prediction algorithm
  - Disparity map
  - Normalized depth map after median scaling
  - Object masks from Mask-RCNN



A. Mauri et al. A new Evaluation Approach for Deep Learning-based Monocular Depth Estimation Methods. ICSCC 2021. 03/07/2021.

Experimental Results under KITTI dataset

- **Monocular Depth Estimation Methods Evaluation: Monotdepth2 vs BTS:**
  - Depth errors for distance rangers of 10m and up to 80m

	RE		SRE		RMSE		logRMSE		a_1		a_2		a_3	
Distance ranges	MD2	BTS	MD2	BTS	MD2	BTS	MD2	BTS	MD2	BTS	MD2	BTS	MD2	BTS
$0-80m$	0.115	**0.060**	0.882	**0.249**	4.701	**2.798**	0.190	**0.096**	0.879	**0.955**	0.961	**0.993**	0.982	**0.998**
$0-10m$	0.102	**0.071**	0.503	**0.188**	1.489	**0.991**	0.141	**0.106**	0.929	**0.959**	0.979	**0.988**	0.99	**0.994**
$10-20m$	0.116	**0.088**	0.845	**0.395**	3.035	**2.198**	0.18	**0.149**	0.891	**0.924**	0.96	**0.971**	0.979	**0.985**
$20-30m$	0.168	**0.13**	1.866	**1.055**	6.208	**4.745**	0.261	**0.229**	0.773	**0.836**	0.916	**0.934**	0.957	**0.964**
$30-40m$	0.196	**0.16**	2.788	**1.945**	9.11	**7.476**	0.307	**0.279**	0.694	**0.764**	0.886	**0.906**	0.942	**0.947**
$40-50m$	0.209	**0.174**	3.504	**2.64**	11.682	**10.008**	0.318	**0.298**	0.641	**0.725**	0.865	**0.889**	**0.943**	0.941
$50-60m$	0.221	**0.19**	4.394	**3.739**	14.252	**12.852**	0.332	**0.326**	0.583	**0.675**	0.857	**0.868**	**0.927**	0.922
$60-70m$	0.212	**0.201**	4.657	**4.584**	15.855	**15.585**	**0.325**	0.334	0.609	**0.619**	0.854	**0.856**	**0.93**	0.923
$70-80m$	**0.181**	0.214	**4.34**	5.454	**15.8**	18.219	**0.284**	0.333	**0.652**	0.548	**0.873**	0.843	**0.945**	0.925

A. Mauri et al. A new Evaluation Approach for Deep Learning-based Monocular Depth Estimation Methods. ICSCC 2021. 03/07/2021.

Experimental Results Under KITTI dataset

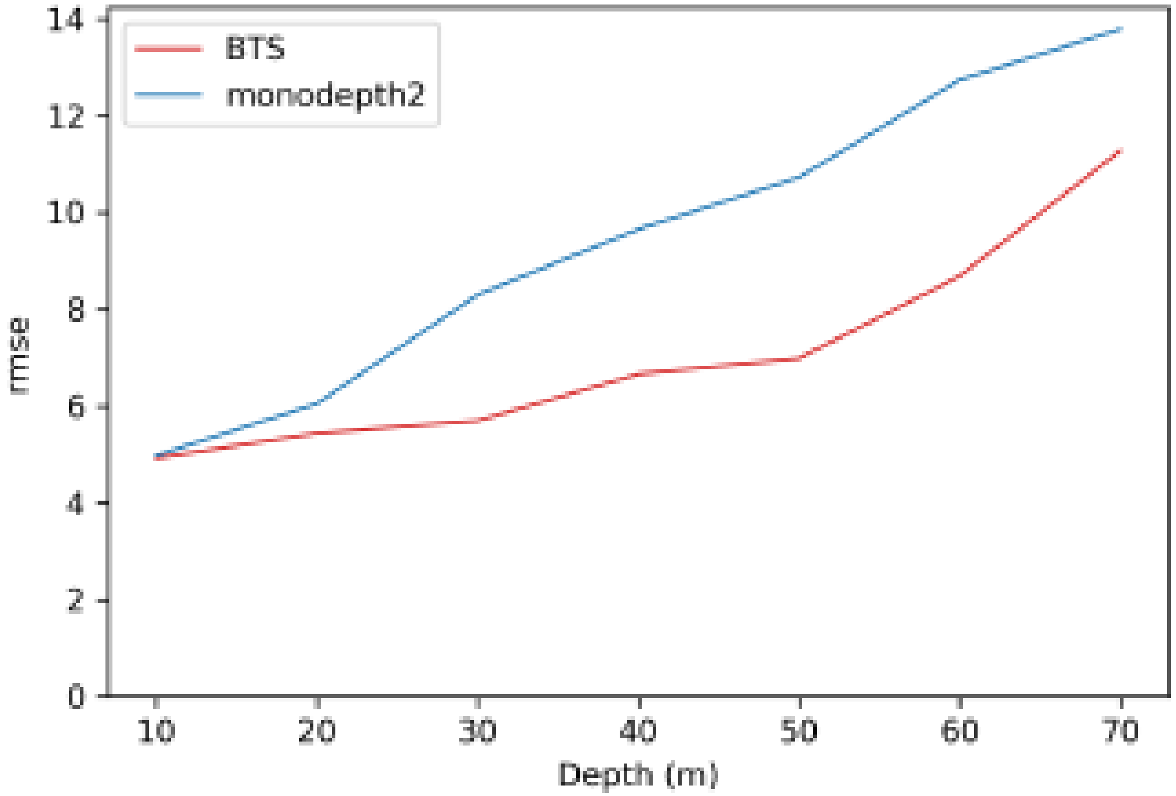
- **Monocular Depth Estimation Methods Evaluation: Monotdepth2 (MD2) vs BTS:**
  - Depth errors computed for the object classes with enough instance in test split distance rangers of 10m and up to 80m

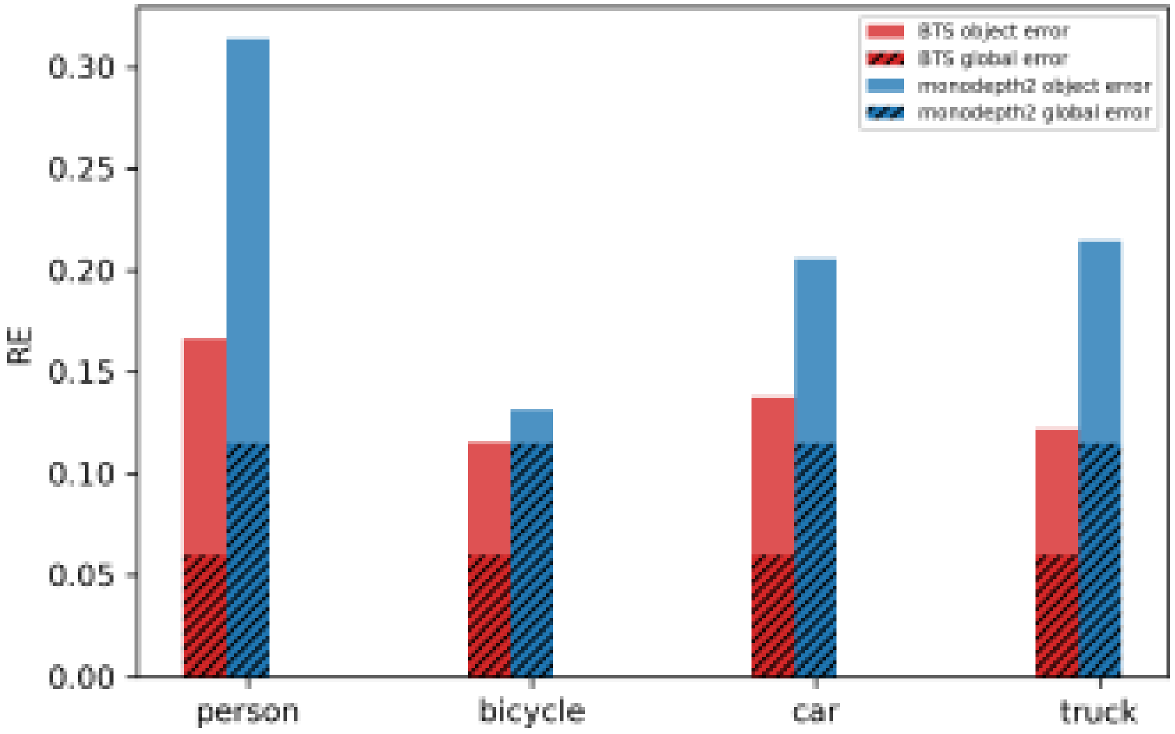
	RE		SRE		RMSE		logRMSE		a_1		a_2		a_3	
Object class	MD2	BTS	MD2	BTS	MD2	BTS	MD2	BTS	MD2	BTS	MD2	BTS	MD2	BTS
Person	0.314	**0.166**	5.721	**1.786**	8.43	**5.892**	0.326	**0.253**	0.601	**0.772**	0.829	**0.894**	0.92	**0.947**
Bicycle	0.131	**0.116**	0.517	**0.467**	2.81	**2.669**	0.172	**0.163**	0.829	**0.839**	**0.964**	0.962	0.993	**0.994**
Car	0.206	**0.137**	3.132	**1.491**	7.924	**6.052**	0.271	**0.223**	0.773	**0.838**	0.883	**0.922**	0.938	**0.955**
Truck	0.215	**0.122**	2.769	**0.826**	6.978	**4.523**	0.259	**0.177**	0.694	**0.854**	0.903	**0.969**	0.964	**0.985**

A. Mauri et al. A new Evaluation Approach for Deep Learning-based Monocular Depth Estimation Methods. ICSCC 2021. 03/07/2021.

Experimental Results under KITTI dataset

- **Quantitative RMSE and RE Results for the car object class over distance ranges on the KITTI dataset.**

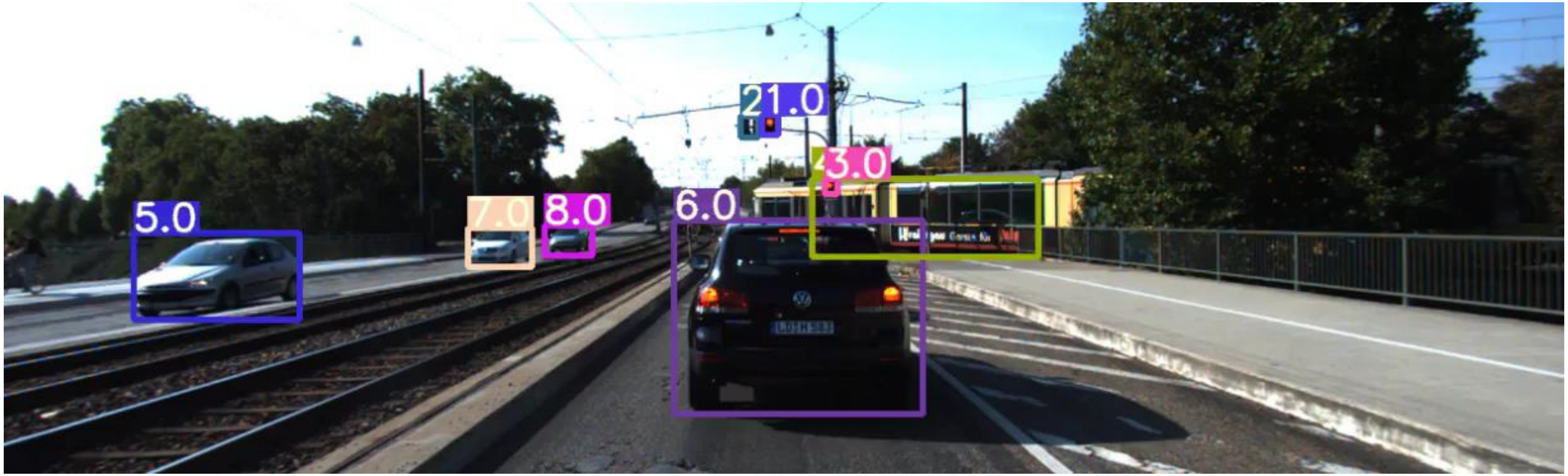




A. Mauri et al. A new Evaluation Approach for Deep Learning-based Monocular Depth Estimation Methods. ICSCC 2021. 03/07/2021.

Experimental results

- **Object Detection**

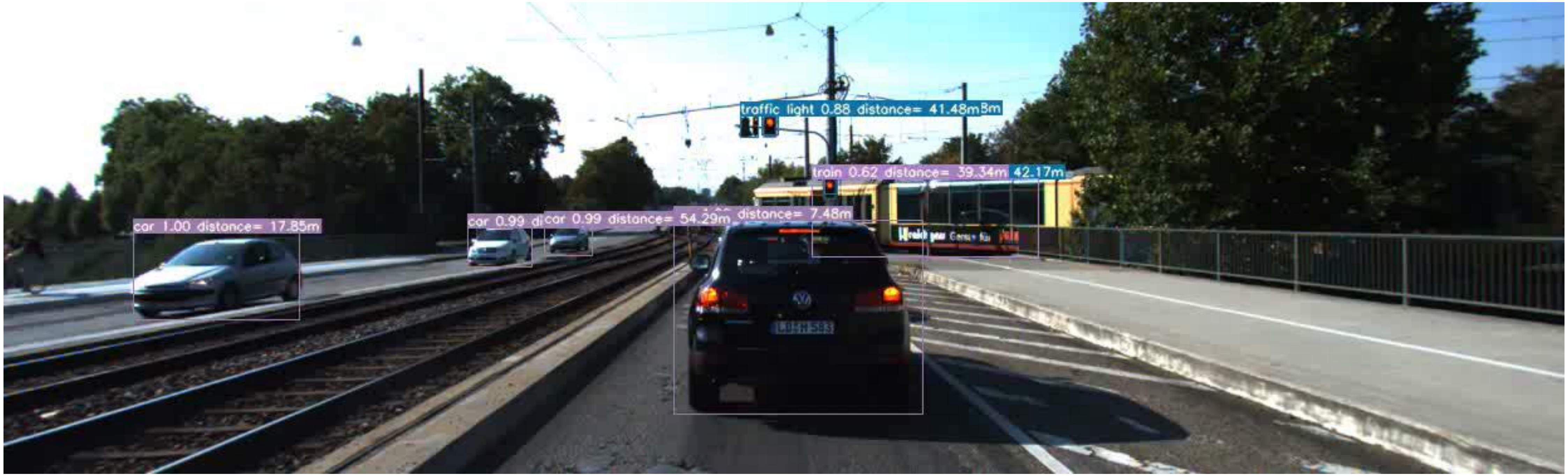


A. Mauri et al. A new Evaluation Approach for Deep Learning-based Monocular Depth Estimation Methods. ICSCC 2021. 03/07/2021.

SEGULA TECHNOLOGIES

Experimental results

- **Distance Estimation of in Traffic Environment under KITTI**



A. Mauri et al. A new Evaluation Approach for Deep Learning-based Monocular Depth Estimation Methods. ICSCC 2021. 03/07/2021.

Experimental results

Depth Estimation Evaluation Methods: Evaluation

A. Mauri et al. A new Evaluation Approach for Deep Learning-based Monocular Depth Estimation Methods. ICSCC 2021. 03/07/2021.

Conclusion and Future Work

- In this work we presented
  - An evaluation of State-of-The-Art for both:
    - Stereo and Mono depth estimation methods

- Our work also showed that
  - BTS is more accurate than Monodepth2
  - GWCnet outperforms PSMNet
  - Stereoscopic methods have a greater accuracy than monocular-based methods

- Future work
  - Evaluate depth estimation algorithms on Rail Environment
  - Acquire our own Railway Dataset

A. Mauri et al. A new Evaluation Approach for Deep Learning-based Monocular Depth Estimation Methods. ICSCC 2021. 03/07/2021.