Supplementary Materials for CabNIR: A Benchmark for In-Vehicle Infrared Monocular Depth Estimation

Ugo Leone Cavalcanti[1] Matteo Poggi[1] Fabio Tosi[1]
Valerio Cambareri[2] Vladimir Zlokolica[2] Stefano Mattoccia[1]

[1]Department of Computer Science and Engineering, University of Bologna, Italy
[2]Sony Depthsensing Solutions, Brussels, Belgium

Project page: https://cabnir.github.io/

1. Responsibility to Human Subjects

Since our dataset aims at in-cabin monitoring, it necessarily features the presence of people whose faces are visible. Indeed, although faces represent Personally Identifiable Information (PII), we cannot hide or blur these details to avoid altering the depth estimation process. As such, all the subjects involved in our acquisitions have been made perfectly aware of the information stored (NIR images and depth maps) and purposes. Each participant agreed to sign an explicit consent form.

Furthermore, our dataset has been approved by our institution's Institutional Review Board (IRB). According to the IRB guidance, CabNIR will be available only to registered users: they shall provide a short overview of their research goal and use the data only for scientific purposes, targeting in-cabin monitoring through depth estimation.

2. Data Description

In this section, we describe how the dataset is structured (2.1) and the car models used for data acquisition (2.2).

2.1. Naming and Structure

We now describe how file names and the directory tree are constructed. Frames related to each recording session are contained in a directory named as the following scheme:

[model_name]_[version]_[sequence]_[Day|Night]

- [model_name] is the car model.
- [version] is an identification letter used to distinguish between different version of the same car model.
- [sequence] is a progressive number to identify distinct recordings of the same cabin (e.g. a different passengers configuration).
- [Day|Night] indicates whether the recording session has been done in daytime or at night.

2.2. Recordings

The dataset comprises 47 scenes featuring 45 people and 36 different cabins. The complete list of the models we employed is in Tab. 1. In some of the scenes we used different version of the same model - e.g. a diverse upholstery, interior configuration or accessories. The use of different versions is denoted by the [version_ID] letter. However, there is also a collection of scenes made with the same cabin, this is indicated by the same [version_ID] letter and a different [sequence_ID] number. As an example: in Yaris_A_1_Night and Yaris_A_2_Night we have the same cabin and camera pose, but in the first sequence the driver is with a passenger, while in the second sequence he is alone.

Sequence Name	Model	# of Seats	Ceiling	Front	Back	Camera Pose	Everyday Objs
500_A_1_Night	Fiat 500	4	Glass	Driver Alone	1 Passenger	Low	✓
500_B_1_Day	Fiat 500	4	Soft Top	Driver+Passenger	Empty	High	✓
500_C_1_Night	Fiat 500	4	Soft Top	Driver+Passenger	Empty	High	✓
500_D_1_Night	Fiat 500	4	Hard Top	Driver Alone	Empty	High	✗
A1_A_1_Day	Audi A1	5	Hard Top	Driver Alone	Empty	High	✓
A3_A_1_Day	Audi A3	5	Hard Top	Driver Alone	Empty	High	✗
A3_A_2_Day	Audi A3	5	Hard Top	Driver+Passenger	Empty	Low	✓
A3_B_1_Night	Audi A3	5	Hard Top	Driver Alone	Empty	High	✗
Beetle_A_1_Day	Volkswagen Beetle	4	Soft Top	Driver+Passenger	Empty	Low	✓
Beetle_A_2_Day	Volkswagen Beetle	4	Soft Top	Driver+Passenger	2 Passenger	Low	✓
Beetle_A_3_Day	Volkswagen Beetle	4	Soft Top	Driver Alone	1 Passenger	Low	✓
C3_A_1_Night	Citroen C3	5	Glass+Fabric	Driver+Passenger	1 Passenger	Low	✗
C3_A_2_Night	Citroen C3	5	Glass+Fabric	Driver+Passenger	Empty	Low	✗
CLA_A_1_Day	Mercedes CLA	5	Hard Top	Driver+Passenger	Empty	Low	✗
Empty	Fiat 500	4	Soft Top	Empty	Empty	High	✓
Fortwo_A_1_Night	Smart Fortwo	2	Soft Top	Driver Alone	-	Low	✗
Fortwo_B_1_Night	Smart Forwo	2	Glass	Driver+Passenger	-	High	✓
Fortwo_C_1_Night	Smart Forwo	2	Hard Top	Driver+Passenger	-	Low	✓
GX3_A_1_Day	Mazda GX3	5	Hard Top	Driver Alone	Empty	Low	✓
Golf_A_1_Night	Volkswagen Golf	5	Galss	Driver+Passenger	Empty	Low	✓
Ibiza_A_1_Day	Seat Ibiza	5	Hard Top	Driver+Passenger	1 Passenger	Low	✓
Jimny_A_1_Day	Suzuki Jimny	3	Hard Top	Driver+Passenger	Empty	High	✗
Mito_A_1_Day	Alfa Romeo Mito	4	Hard Top	Driver+Passenger	Empty	Low	✓
Model3_A_1_Night	Tesla Model 3	5	Glass	Driver	Empty	Low	✓
Model3_A_2_Night	Tesla Model 3	5	Glass	Driver+Passenger	Empty	Low	✓
Panda_A_1_Day	Fiat Panda	4	Hard Top	Driver+Passenger	Empty	High	✗
Panda_B_1_Day	Fiat Panda	5	Hard Top	Driver Alone	Empty	High	✗
Panda_C_1_Night	Fiat Panda	5	Hard Top	Driver Alone	Empty	Low	✗
Panda_D_1_Day	Fiat Panda	4	Hard Top	Driver Alone	Empty	Low	✗
Panda_E_1_Night	Fiat Panda	5	Hard Top	Driver+Passenger	Empty	Low	✗
Panda_E_2_Night	Fiat Panda	5	Hard Top	Driver Alone	Empty	Low	✗
Panda_F_1_Night	Fiat Panda	5	Hard Top	Driver+Passenger	Empty	Low	✗
Panda_F_2_Night	Fiat Panda	5	Hard Top	Driver+Passenger	1 Passenger	Low	✗
Polo_A_1_Night	Volkswagen Polo	5	Hard Top	Driver Alone	Empty	High	✓
Puma_A_1_Night	Ford Puma	5	Hard Top	Driver Alone	Empty	Low	✓
RS3_A_1_Night	Audi RS3	5	Glass	Driver+Passenger	1 Passenger	High	✗
Up_A_1_Day	Volkswagen Up	4	Hard Top	Driver Alone	Empty	High	✗
Up_B_1_Night	Volkswagen Up	4	Hard Top	Driver+Passenger	Empty	Low	✗
V60_A_1_Night	Volvo V60	5	Hard Top	Driver Alone	Empty	Low	✗
X2_A_1_Night	BMW X2	5	Hard Top	Driver Alone	Empty	Low	✓
Yaris_A_1_Night	Toyota Yaris	5	Hard Top	Driver+Passenger	Empty	Low	✓
Yaris_A_2_Night	Toyota Yaris	5	Hard Top	Driver Alone	Empty	Low	✓
Yaris_B_1_Night	Toyota Yaris	5	Hard Top	Driver Alone	Empty	High	✗
Yaris_C_1_Night	Toyota Yaris	5	Hard Top	Driver Alone	Empty	Low	✗
Yaris_D_1_Night	Toyota Yaris	5	Hard Top	Driver Alone	Empty	Low	✗
Yaris_E_1_Day	Toyota Yaris	5	Hard Top	Driver+Passenger	Empty	Low	✓
Yaris_E_2_Day	Toyota Yaris	5	Hard Top	Driver+Passenger	1 Passenger	Low	✓

Table 1. **CabNIR-Sequences Description**

	Validation Split						Test Split					
Model	**AbsRel**↓	**RMSE**↓	**MAE**↓	$\delta_{1.10}$ ↑	$\delta_{1.20}$ ↑	$\delta_{1.30}$ ↑	**AbsRel**↓	**RMSE**↓	**MAE**↓	$\delta_{1.10}$ ↑	$\delta_{1.20}$ ↑	$\delta_{1.30}$ ↑
MiDaS [4]	0.435	0.234	0.171	0.227	0.433	0.587	0.383	0.205	0.154	0.247	0.442	0.572
LeReS [6]	0.507	0.289	0.217	0.180	0.332	0.455	0.432	0.237	0.165	0.253	0.451	0.584
DPT [3]	**0.291**	**0.164**	**0.128**	**0.268**	**0.480**	**0.642**	**0.272**	**0.149**	**0.115**	**0.278**	**0.508**	**0.682**
OmniData [1]	0.370	0.227	0.169	0.208	0.406	0.564	0.455	0.233	0.178	0.187	0.358	0.506
Depth Anything [5]	0.312	0.187	0.143	0.228	0.462	0.633	0.329	0.178	0.133	0.251	0.500	0.654

Table 2. **Zero-shot affine-invariant networks.** We report the results achieved by several existing networks trained on large-scale datasets. Results on validation (left) and test (right) splits.

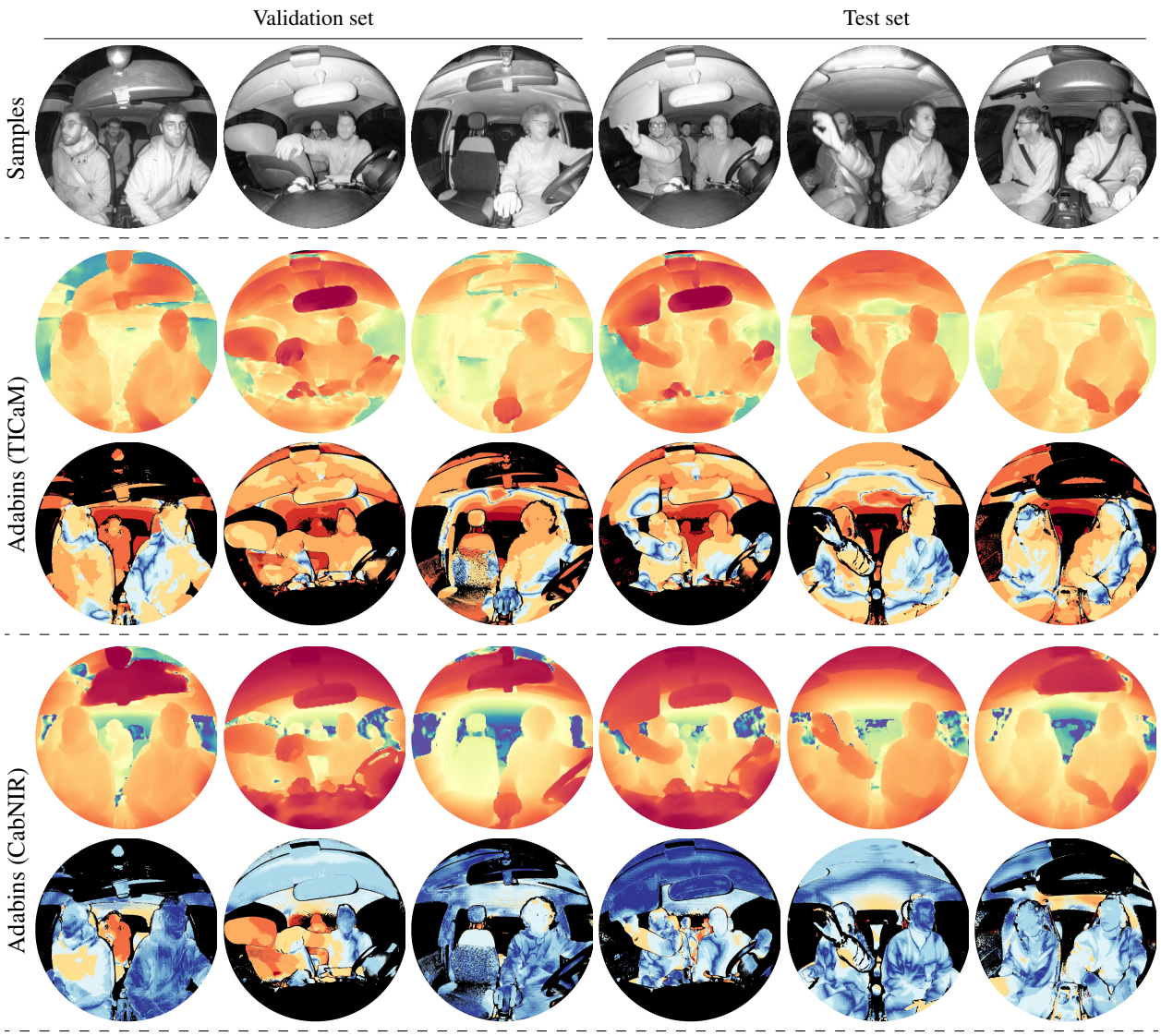

Figure 1. **Qualitative results on CabNIR validation and test split – Adabins.** Predictions and error maps by models trained on TICaM [2] or our training set.

3. Zero-Shot Results

Tab. 2 shows the results achieved by running the pre-trained models – MiDaS [4], LeReS [6], DPT [3] and OmniData [1] and Depth Anything [5] – directly on CabNIR without any fine-tuning. Despite being trained on millions of images, the networks fail to generalize to the very different setting features in our dataset, both in terms of image modality (NIR) and camera setting (wide-angle). Among the evaluated models, DPT demonstrates better generalization to our domain, achieving an average error close to 10cm.

4. Qualitative Results

We conclude by reporting some qualitative results. Fig. 1 shows a comparison between the predictions by Adabins trained on TICaM and CabNIR respectively. Despite the better results yielded by the latter, we can appreciate the high errors in the presence of rare conditions – e.g., the slanted seat in column 2, or the raised hand in column 7. Fig. 2 concludes this experiment by qualitatively comparing the original and fine-tuned DPT models. However, despite the significant improvement achieved with the fine-tuning, DPT still fails in the presence of the slanted seat in column 2.

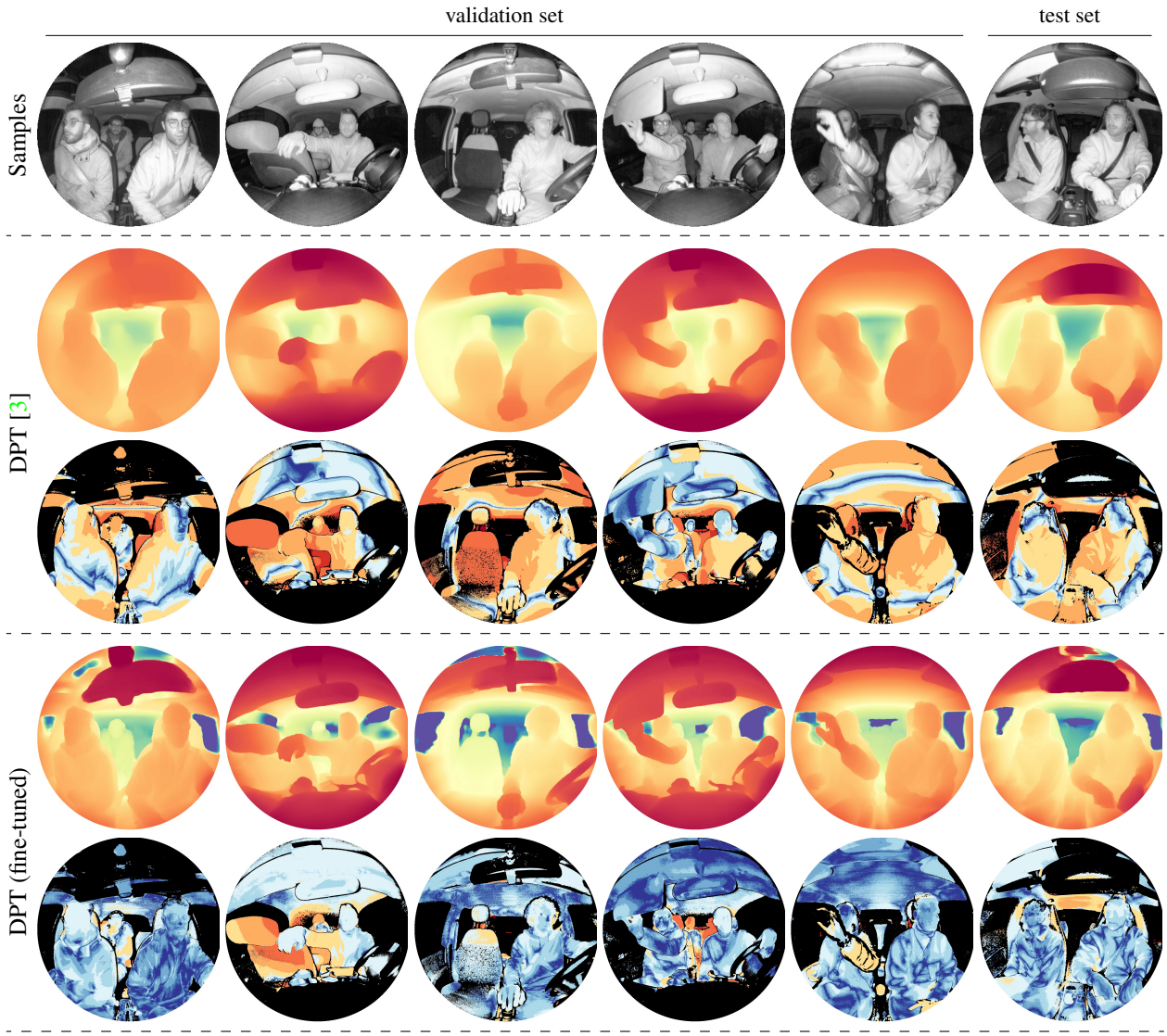

Figure 2. **Qualitative results on CabNIR validation and test split – DPT.** Predictions and error maps by the original model or the one fine-tuned on our training set.

References

[1] Ainaz Eftekhar, Alexander Sax, Jitendra Malik, and Amir Zamir. Omnidata: A scalable pipeline for making multi-task mid-level vision datasets from 3d scans. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10786–10796, 2021. 2, 3

[2] Jigyasa Singh Katrolia, Ahmed El-Sherif, Hartmut Feld, Bruno Mirbach, Jason R. Rambach, and Didier Stricker. Ticam: A time-of-flight in-car cabin monitoring dataset. In *32nd British Machine Vision Conference 2021, BMVC 2021, Online, November 22-25, 2021*, page 277. BMVA Press, 2021. 3

[3] René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. Vision transformers for dense prediction. *ICCV*, 2021. 2, 3, 4

[4] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(3), 2022. 2, 3

[5] Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything: Unleashing the power of large-scale unlabeled data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10371–10381, 2024. 2, 3

[6] Wei Yin, Jianming Zhang, Oliver Wang, Simon Niklaus, Long Mai, Simon Chen, and Chunhua Shen. Learning to recover 3d scene shape from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 204–213, 2021. 2, 3