

000
001
002
003
004
005
006
007
008
009
010
011
012
013
014
015
016
017
018
019
020
021
022
023
024
025
026
027
028
029
030
031
032
033
034
035
036
037
038
039
040
041
042
043
044
045
046
047
048
049
050
051
052
053054
055
056
057
058
059
060
061
062
063
064
065
066
067
068
069
070
071
072
073
074
075
076
077
078
079
080
081
082
083
084
085
086
087
088
089
090
091
092
093
094
095
096
097
098
099
100
101
102
103
104
105
106
107

Supplementary Material: Matching-space Stereo Networks for Cross-domain Generalization

Anonymous 3DV submission

Paper ID 36

In this supplementary document, we present additional qualitative results on domain generalization (i.e., *sf-all* → *others*), which are omitted in the main paper due to space limits.

1. Qualitative Results from sf-all → KT15

In Fig. 1 we provide more qualitative results on KITTI 2015 (KT15) [4] for the networks trained on *sf-all* [3]. Every two rows correspond to an example from the KT15 training set. Specifically, Fig. 1(a) shows the reference image and ground truth, Fig. 1(b) to (e) show the disparity (top) and error (bottom) maps obtained by baseline GCNet [2], our MS-GCNet, baseline PSMNet [1] and our MS-PSMNet, respectively. Please note the bad3-all rates are superimposed on the error maps.

2. Qualitative Results from sf-all → MB

Fig. 2 shows additional qualitative results on Middlebury 2014 (MB) [5] for the networks trained on *sf-all*. Each row corresponds to an example from the MB training set. Fig. 2(a) shows the reference image, Fig. 2(b) to (e) are the disparity maps obtained by baseline GCNet [2], our MS-GCNet, baseline PSMNet [1] and our MS-PSMNet, respectively. Bad2-noc rates are superimposed on the disparity maps.

3. Qualitative Results from sf-all → KT Raw

We show more qualitative results on KITTI (KT) raw sequence *2011_10_03_drive_0034_sync*, for MS-GCNet (Fig. 3) and MS-PSMNet (Fig. 4) trained on *sf-all*. Specifically, Fig. 3(a) is the input left frame, and Fig. 3(b) and (c) are the disparity maps estimated by baseline GCNet [2] and our MS-GCNet, respectively. Fig 4 provides the results for baseline PSMNet [1] and our MS-PSMNet. Please note there is no ground truth for KT raw sequences, so the error rates cannot be calculated. Still, comparing the disparity maps in (b) and (c), ours tend to predict more reliable and smooth disparities rather than the noisy and bumpy ones by

the baselines. For more examples, please see the *videos* attached in this supplementary material for our MS-GCNet and MS-PSMNet. Each frame of the videos shows the input left image (top), the disparity map estimated by the baseline (middle), and the disparity map by our MS counterpart (bottom).

References

- [1] J.-R. Chang and Y.-S. Chen. Pyramid stereo matching network. In *CVPR*, pages 5410–5418, 2018. 1, 2, 3, 5
- [2] A. Kendall, H. Martirosyan, S. Dasgupta, P. Henry, R. Kennedy, A. Bachrach, and A. Bry. End-to-end learning of geometry and context for deep stereo regression. In *ICCV*, pages 66–75, 2017. 1, 2, 3, 4
- [3] N. Mayer, E. Ilg, P. Hausser, P. Fischer, D. Cremers, A. Dosovitskiy, and T. Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In *CVPR*, pages 4040–4048, 2016. 1
- [4] M. Menze and A. Geiger. Object scene flow for autonomous vehicles. In *CVPR*, pages 3061–3070, 2015. 1, 2
- [5] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nešić, X. Wang, and P. Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In *German Conference on Pattern Recognition*, pages 31–42, 2014. 1, 3

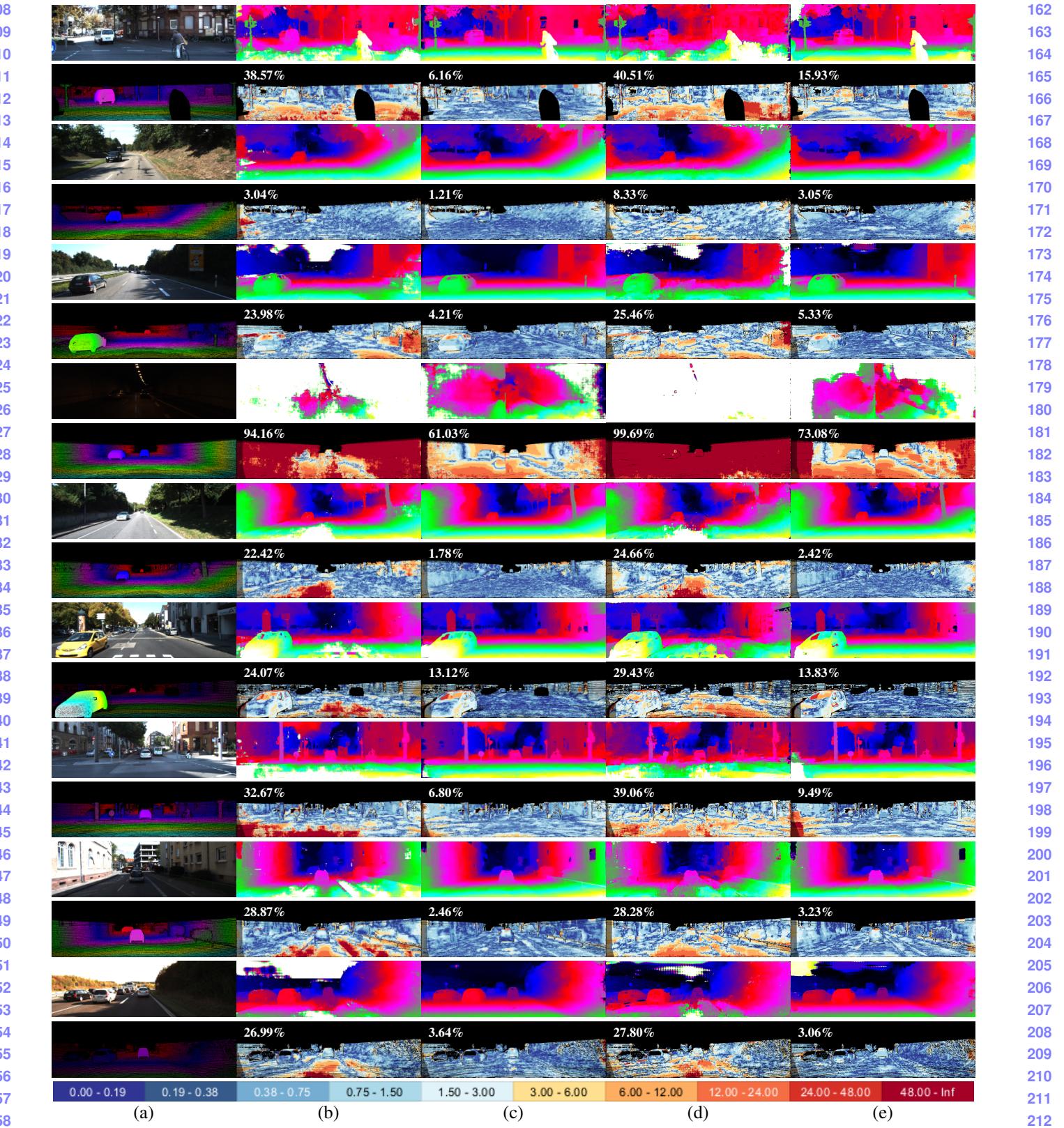


Figure 1: Qualitative results on KITTI 2015 [4] for networks trained on *sf-all*. Every two rows correspond to one example in KITTI 2015 training set. (a) shows reference image and ground truth, (b) to (e) disparity (top) and error (bottom) maps obtained with GCNet [2], MS-GCNet, PSMNet [1] and MS-PSMNet, respectively. Bad3-all rates are superimposed on the error maps.

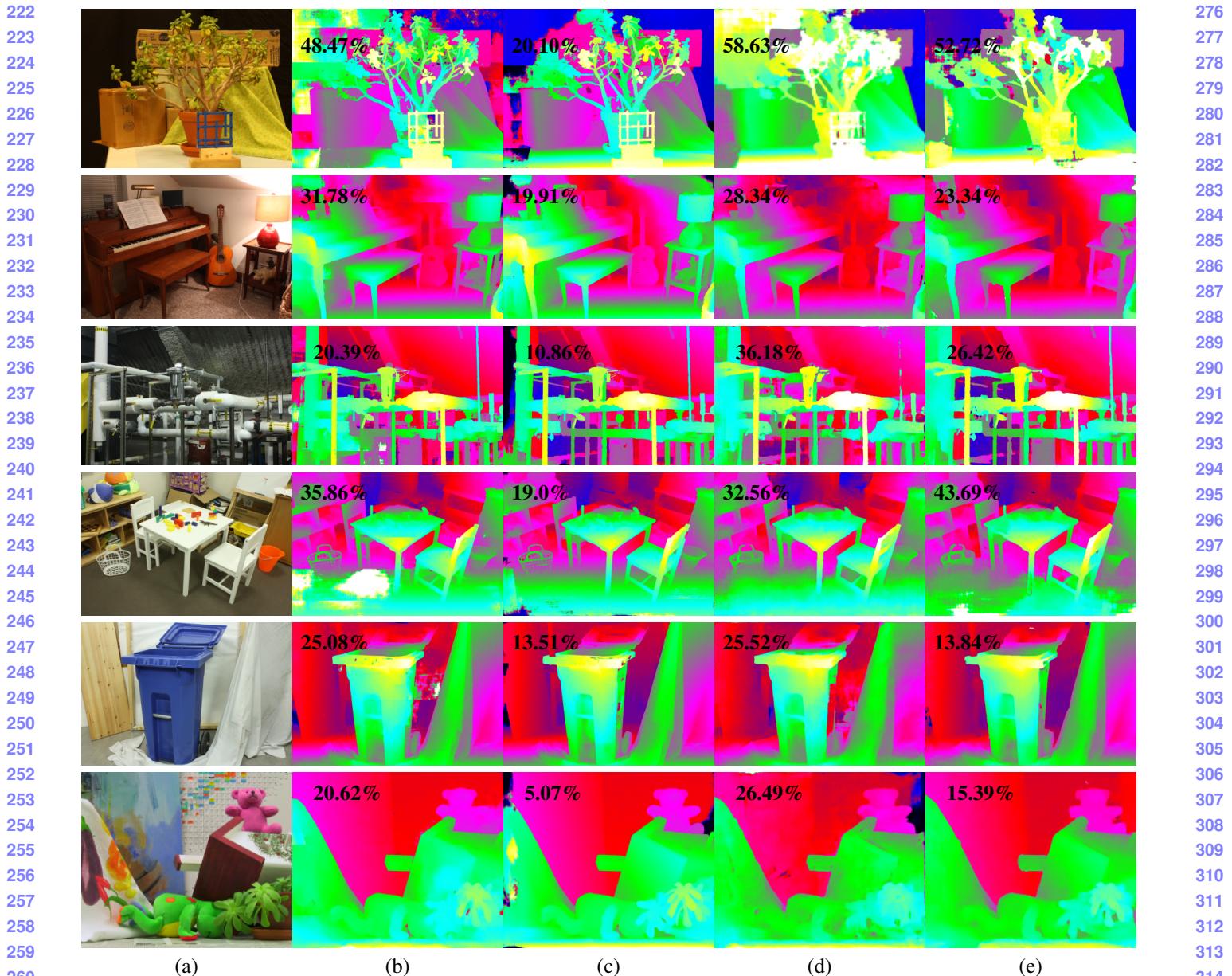
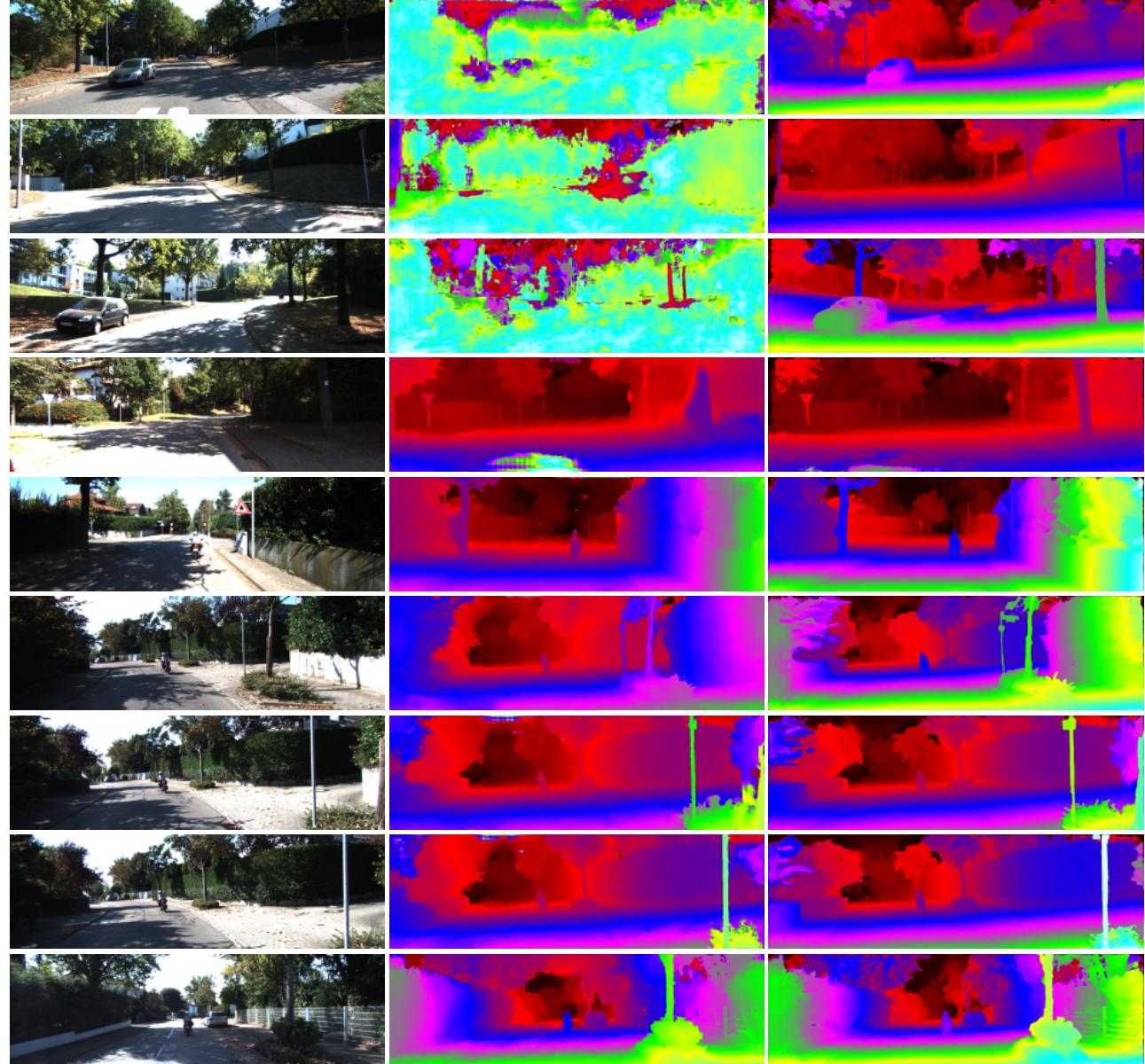
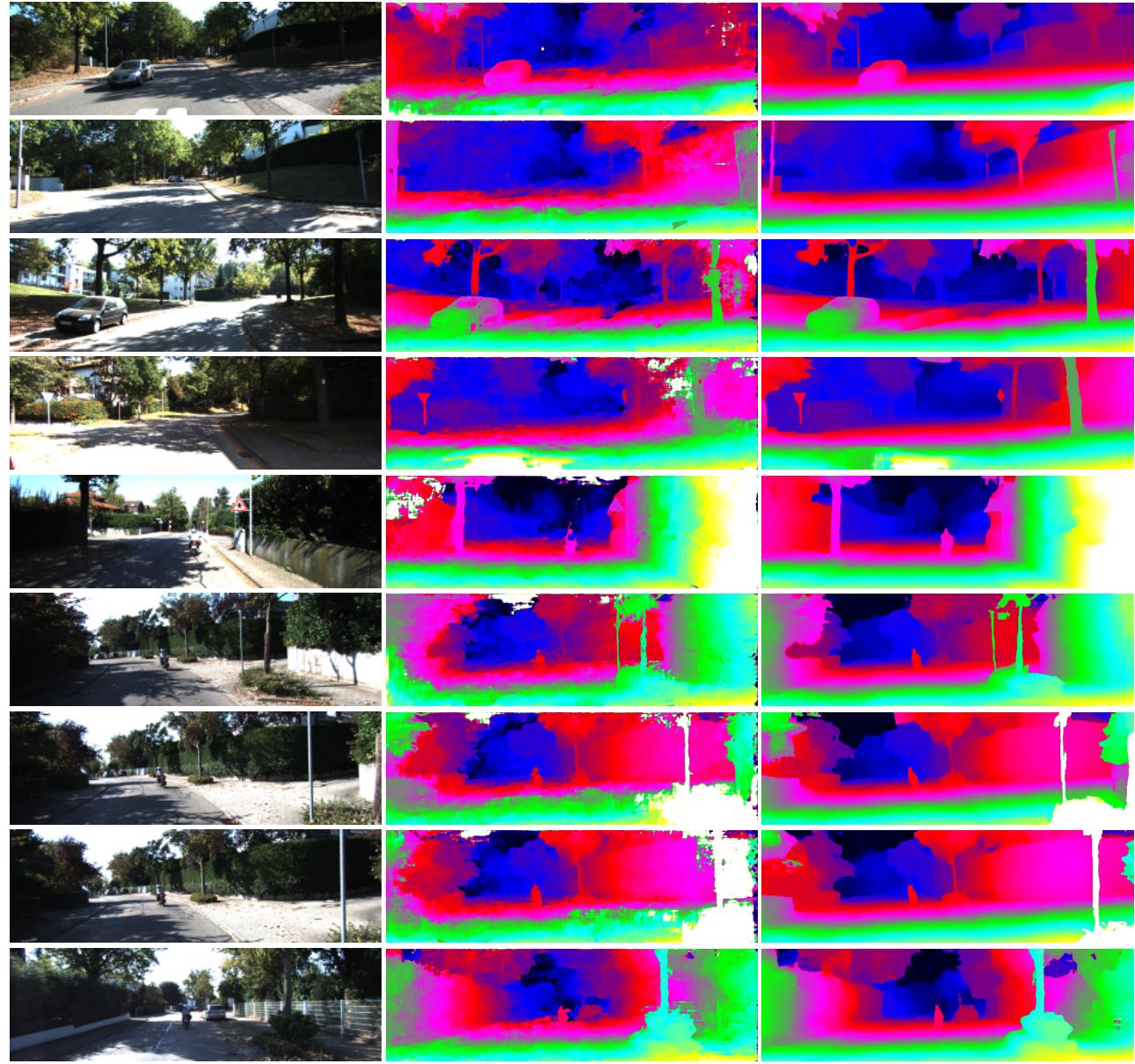
216
217
218
219
220
221

Figure 2: Qualitative results on Middlebury 2014 [5] for networks trained on *sf-all*. Column (a) shows reference image, (b) to (e) disparity maps obtained with GCNet [2], MS-GCNet, PSMNet [1] and MS-PSMNet, respectively. Bad2-noc rates are superimposed on the disparity maps.



369 Figure 3: Qualitative results on KITTI raw sequence 2011_10_03_drive_0034_sync, for networks trained on *sf-all*. (a) Left input frame. (b)
370 Disparity map estimated by GCNet [2]. (c) Disparity map estimated by our MS-GCNet. Both networks are trained on the same synthetic
371 data *sf-all*.



(a)

(b)

(c)

Figure 4: Qualitative results on KITTI raw sequence *2011_10_03_drive_0034_sync*, for networks trained on *sf-all*. (a) Left input frame. (b) Disparity map estimated by PSMNet [1]. (c) Disparity map estimated by our MS-PSMNet. Both networks are trained on the same synthetic data *sf-all*.