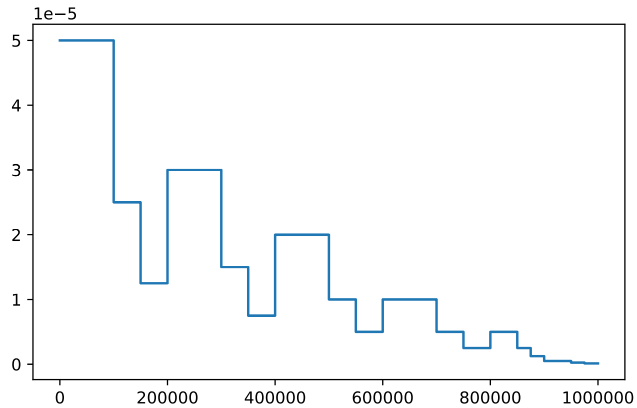
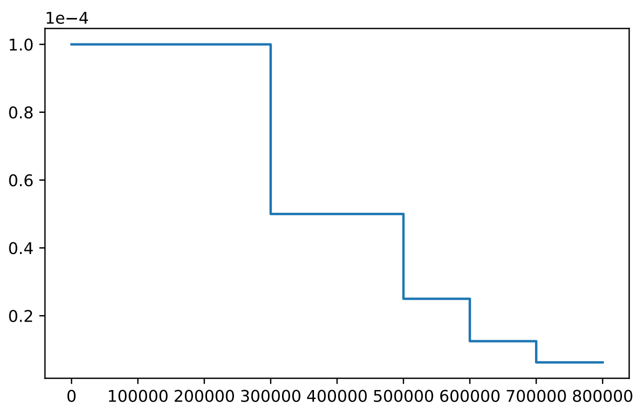
Appendix A. More Implementation Details

Training Schedule. When fine-tuning on Sintel, we use a longer schedule (see Fig. 11(a)) referring to the cyclic learning rate proposed by PWC-Net+ [31]. When training the second stage, we follow again the same schedule as the first stage for all datasets except that it is shorter on FlyingChairs (see Fig. 11(b)). For submission to the test set, we train on the whole training set and reduce randomness by averaging 3 independent runs due to the huge variance.



(a) Schedule for fine-tuning on Sintel.



(b) A shorter schedule for the second stage on FlyingChairs.

Figure 11. **Learning rate schedules.**

Data Augmentation. We implement geometric and chromatic augmentations referring to the implementation of FlowNet [8] and IRR-PWC [13]. Details about the sampling ranges for each training stage are provided in Table 6 (for geometric augmentations) and Table 7 (for chromatic augmentations). We use the same augmentations on FlyingThings3D as FlyingChairs. We finally apply a random crop (within valid areas) using a size of 448×320 on FlyingChairs, 768×384 on FlyingThings3D, 768×320 on Sintel, and 896×320 on KITTI. To avoid out-of-bound areas after cropping, we compute the minimum degree of zoom that forces the existence of a valid crop.

Geometric Aug.	Chairs	Sintel	KITTI
Horizontal Flip	0.5	0.5	0.5
Squeeze	0.9	0.9	0.95
Translation	0.1	0.1	0.05
Rel. Translation	0.025	0.025	0.0125
Rotation	$17°$	$17°$	$5°$
Rel. Rotation	$4.25°$	$4.25°$	$1.25°$
Zoom	$[0.9, 2.0]$	$[0.9, 1.5]$	$[0.95, 1.25]$
Rel. Zoom	0.96	0.96	0.98

Table 6. **Geometric augmentations.**

Chromatic Aug.	Chairs	Sintel	KITTI
Contrast	$[-0.4, 0.8]$	$[-0.4, 0.8]$	$[-0.2, 0.4]$
Brightness	0.1	0.1	0.05
Channel	$[0.8, 1.4]$	$[0.8, 1.4]$	$[0.9, 1.2]$
Saturation	0.5	0.5	0.25
Hue	0.5	0.5	0.1
Noise	0.04	0	0.02

Table 7. **Chromatic augmentations.**

Appendix B. More Visualizations

More visualizations of the learnable occlusion mask and the flow predictions are presented in Fig. 12 and Fig. 13. Note that the learned occlusion masks are relatively vague at the image boundary, since the network cannot learn to mask out-of-bound features that are already zeros. We expect that the estimation results can be further improved if out-of-bound areas are manually regarded as occlusions.

Appendix C. Screenshots on Benchmarks

At the time of submission, MaskFlownet ranks first on the MPI Sintel benchmark on both clean pass (see Fig. 14) and final pass (see Fig. 15). Note that the top entry (ScopeFlow) at the time of screenshot (Nov. 23th, 2019) on the final pass is a new anonymous submission, with a relatively poor performance on the clean pass. Remarkably, MaskFlownet outperforms the previous top entry on the clean pass (MR-Flow [38]) that uses the rigidity assumption while being very slow, as well as the previous top entry on the final pass (SelFlow [18]) that uses multi-frame inputs.

On the KITTI 2012 and 2015 benchmarks, MaskFlownet surpasses all optical flow methods (excluding the anonymous entries) at the time of submission (see Fig. 16 and Fig. 17). Note that the top 3 entries on the KITTI 2015 benchmark are *scene flow* methods that use stereo images and thus not comparable.

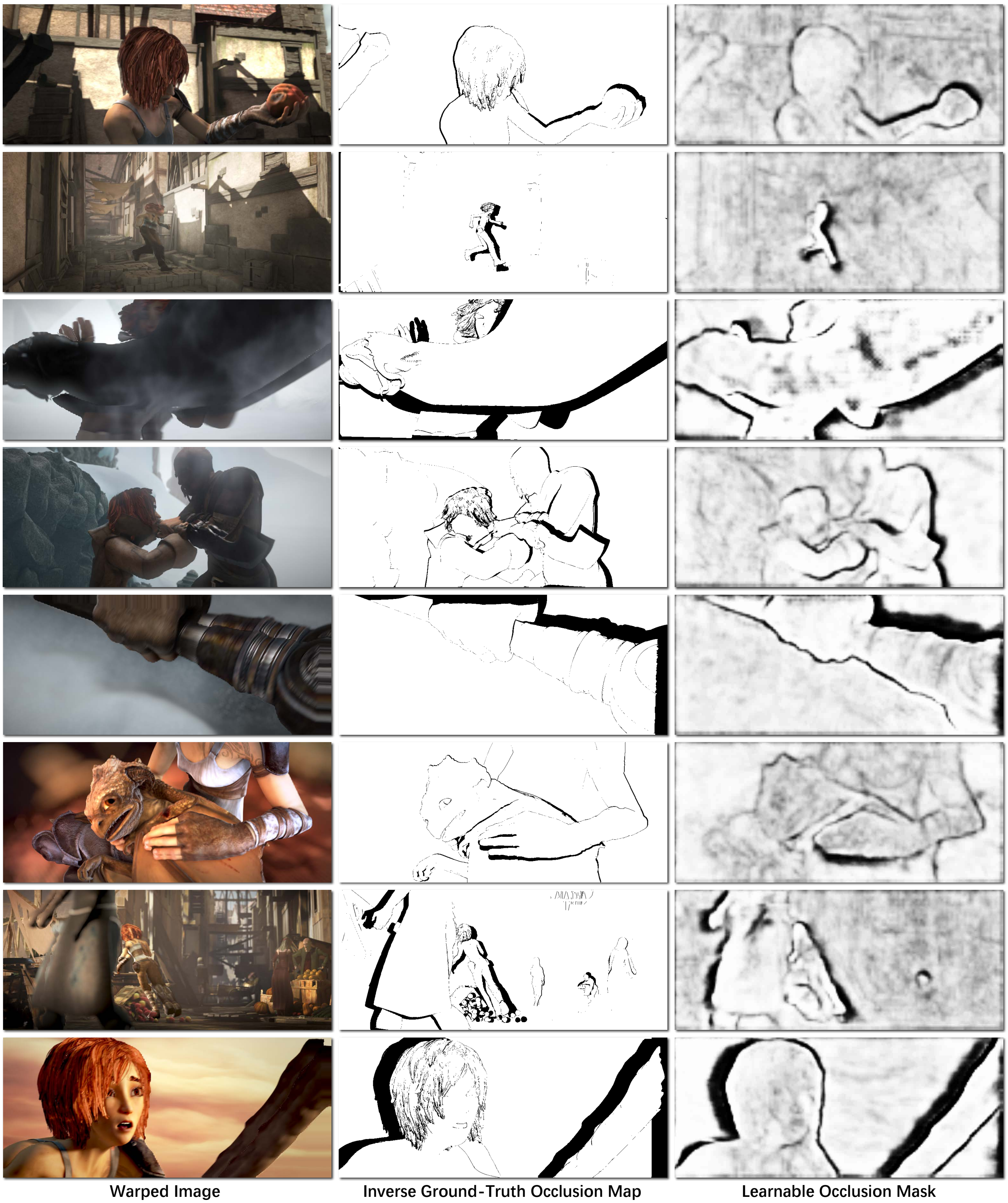


Figure 12. **More visualizations of the learnable occlusion mask.** All samples are chosen from the Sintel training set (final pass). The learnable occlusion masks are expected to (roughly) match the inverse ground-truth occlusion maps, even if they are learned without any explicit supervision.

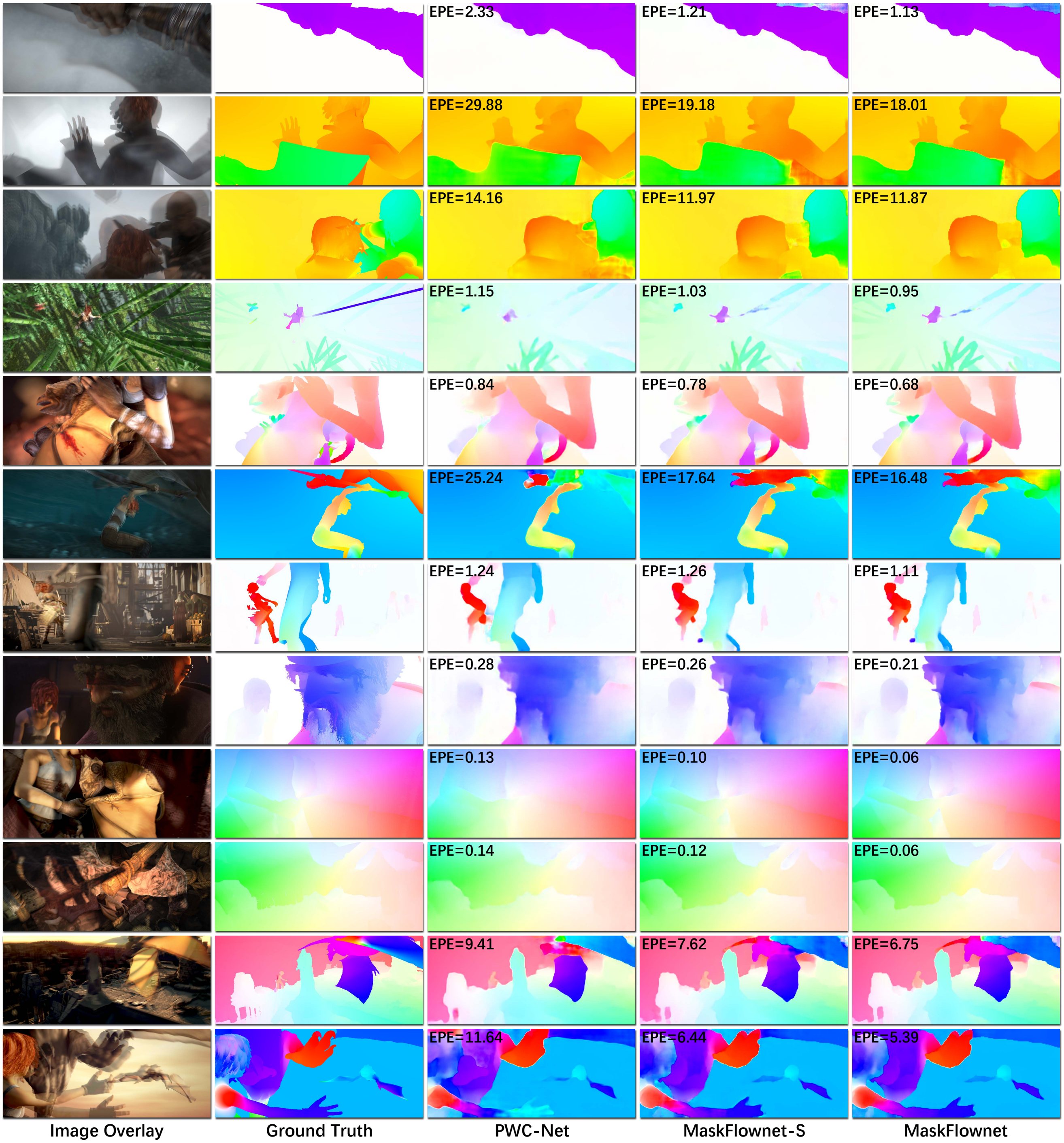
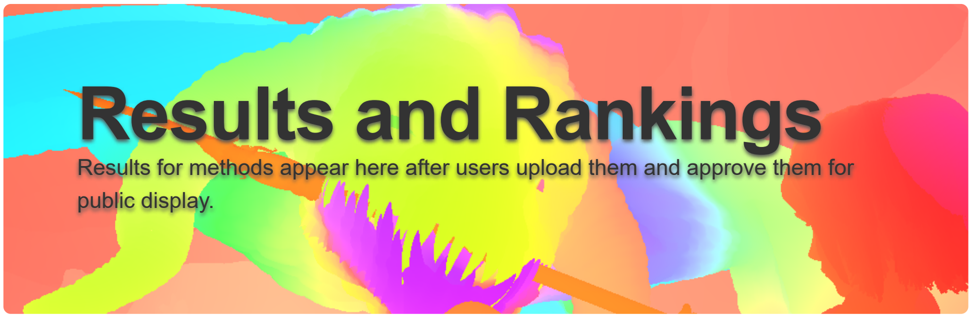


Figure 13. **More visualizations for qualitative comparison among PWC-Net [32], MaskFlownet-S, and MaskFlownet.** All samples are chosen from the Sintel training set (final pass). We replicate PWC-Net using the PyTorch reimplementation [25] that provides a pretrained model of the "PWC-Net_ROB" version [31].

MPI Sintel Dataset About Downloads Results FAQ Contact Signup Login

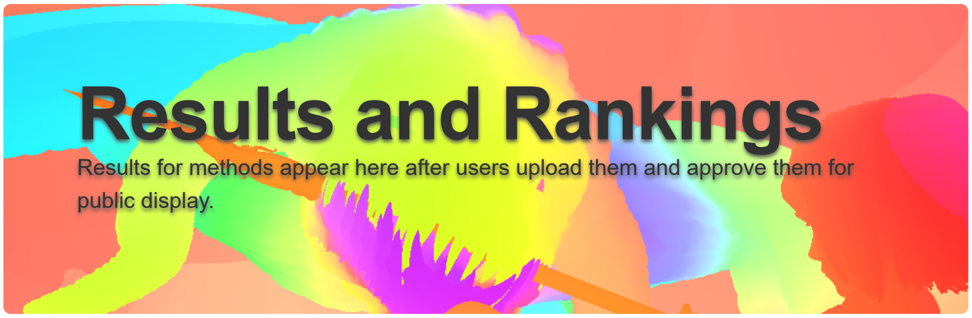


Final Clean

	EPE all	EPE matched	EPE unmatched	d0-10	d10-60	d60-140	s0-10	s10-40	s40+	
GroundTruth [1]	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	Visualize Results
MaskFlownet [2]	2.521	0.989	15.032	2.742	0.908	0.291	0.361	1.285	16.261	Visualize Results
MR-Flow [3]	2.527	0.954	15.365	2.866	0.710	0.420	0.446	1.715	14.826	Visualize Results
ProFlow_ROB [4]	2.709	1.013	16.549	2.843	0.723	0.518	0.485	1.586	16.470	Visualize Results
MaskFlownet-S [5]	2.771	1.077	16.608	2.901	0.996	0.342	0.419	1.404	17.777	Visualize Results
VCN [6]	2.808	1.108	16.682	3.267	0.867	0.418	0.646	1.669	16.302	Visualize Results
ProFlow [7]	2.818	1.027	17.428	2.892	0.751	0.496	0.469	1.626	17.369	Visualize Results
SfM-PM [8]	2.910	1.016	18.357	2.797	0.756	0.479	0.559	1.732	17.431	Visualize Results
FlowFields++ [9]	2.943	0.850	20.027	2.550	0.603	0.403	0.560	1.859	17.401	Visualize Results
LiteFlowNet3 [10]	2.994	1.148	18.077	3.000	0.985	0.498	0.559	1.670	18.302	Visualize Results
FlowFields+ [11]	3.102	0.820	21.718	2.340	0.616	0.373	0.593	1.865	18.549	Visualize Results
DIP-Flow [12]	3.103	0.881	21.227	2.574	0.681	0.419	0.548	1.801	18.979	Visualize Results
PST [13]	3.110	0.942	20.809	2.759	0.664	0.378	0.635	2.069	17.919	Visualize Results

Figure 14. **Screenshot on the MPI Sintel clean pass (printed as PDF).**

MPI Sintel Dataset | About | Downloads | Results | FAQ | Contact | Signup | Login



Final | Clean

	EPE all	EPE matched	EPE unmatched	d0-10	d10-60	d60-140	s0-10	s10-40	s40+	
GroundTruth [1]	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	Visualize Results
ScopeFlow [2]	4.098	1.999	21.214	4.028	1.689	1.180	0.725	2.589	24.477	Visualize Results
MaskFlownet [3]	4.172	2.048	21.494	3.783	1.745	1.310	0.592	2.389	26.253	Visualize Results
SelFlow [4]	4.262	2.040	22.369	4.083	1.715	1.287	0.582	2.343	27.154	Visualize Results
MaskFlownet-S [5]	4.384	2.120	22.840	3.905	1.821	1.359	0.645	2.526	27.429	Visualize Results
VCN [6]	4.404	2.216	22.238	4.381	1.782	1.423	0.955	2.725	25.570	Visualize Results
LiteFlowNet3 [7]	4.448	2.089	23.681	3.873	1.755	1.344	0.754	2.503	27.471	Visualize Results
ContinualFlow_ROB [8]	4.528	2.723	19.248	5.050	2.573	1.713	0.872	3.114	26.063	Visualize Results
MFF [9]	4.566	2.216	23.732	4.664	2.017	1.222	0.893	2.902	26.810	Visualize Results
IRR-PWC [10]	4.579	2.154	24.355	4.165	1.843	1.292	0.709	2.423	28.998	Visualize Results
PWC-Net+ [11]	4.596	2.254	23.696	4.781	2.045	1.234	0.945	2.978	26.620	Visualize Results
PPAC-HD3 [12]	4.599	2.116	24.852	3.521	1.702	1.637	0.617	2.083	30.457	Visualize Results
CompactFlow [13]	4.626	2.099	25.253	4.192	1.825	1.233	0.845	2.677	28.120	Visualize Results

Figure 15. **Screenshot on MPI Sintel final pass (printed as PDF).**

background interpolation as explained in the corresponding header file in the development kit. For each method we show:

- **Out-Noc:** Percentage of erroneous pixels in non-occluded areas
- **Out-All:** Percentage of erroneous pixels in total
- **Avg-Noc:** Average disparity / end-point error in non-occluded areas
- **Avg-All:** Average disparity / end-point error in total
- **Density:** Percentage of pixels for which ground truth has been provided by the method

Note: On 04.11.2013 we have improved the ground truth disparity maps and flow fields leading to slightly improvements for all methods. Please download the stereo/flow dataset with the improved ground truth for training again, if you have downloaded the dataset prior to 04.11.2013. Please consider reporting these new number for all future submissions. Links to last leaderboards before the updates: stereo and flow!

Important Policy Update: As more and more non-published work and re-implementations of existing work is submitted to KITTI, we have established a new policy: from now on, only submissions with significant novelty that are leading to a peer-reviewed paper in a conference or journal are allowed. Minor modifications of existing algorithms or student research projects are not allowed. Such work must be evaluated on a split of the training set. To ensure that our policy is adopted, new users must detail their status, describe their work and specify the targeted venue during registration. Furthermore, we will regularly delete all entries that are 6 months old but are still anonymous or do not have a paper associated with them. For conferences, 6 month is enough to determine if a paper has been accepted and to add the bibliography information. For longer review cycles, you need to resubmit your results.

Additional information used by the methods

- Stereo: Method uses left and right (stereo) images
- Multiview: Method uses more than 2 temporally adjacent images
- Motion stereo: Method uses epipolar geometry for computing optical flow
- Additional training data: Use of additional data sources for training (see details)

Error threshold 3 pixels **Evaluation area** All pixels

	Method	Setting	Code	Out-Noc	Out-All	Avg-Noc	Avg-All	Density	Runtime	Environment	Compare
1	DM-Net-i2		code	**0.00 %**	**0.00 %**	**0.0 px**	0.0 px	0.00 %	0.90 s	1 core @ 2.5 Ghz (C/C++)	
2	Anonym			0.00 %	0.00 %	0.0 px	0.0 px	0.00 %	TBD s	1 core @ 2.5 Ghz (Python)	
3	PPAC-HD3			2.01 %	5.09 %	0.6 px	1.2 px	100.00 %	0.14 s	NVIDIA GTX 1080 Ti	
4	PCF-F			2.07 %	5.45 %	0.6 px	1.2 px	100.00 %	0.08 s	GPU @ 2.5 Ghz (Python)	
5	MaskFlownet			2.07 %	4.82 %	0.6 px	1.1 px	100.00 %	0.06 s	NVIDIA TITAN Xp	
6	HD^3-Flow		code	2.26 %	5.41 %	0.7 px	1.4 px	100.00 %	0.10 s	NVIDIA Pascal Titan XP	
	Z. Yin, T. Darrell and F. Yu: Hierarchical Discrete Distribution Decomposition for Match Density Estimation. CVPR 2019.										
7	MaskFlownet-S			2.29 %	5.24 %	0.6 px	1.1 px	100.00 %	0.03 s	NVIDIA TITAN Xp	
8	PRSM		code	2.46 %	4.23 %	0.7 px	1.0 px	100.00 %	300 s	1 core @ 2.5 Ghz (Matlab + C/C++)	
	C. Vogel, K. Schindler and S. Roth: 3D Scene Flow Estimation with a Piecewise Rigid Scene Model. ijcv 2015.										
9	LiteFlowNet3-S			2.49 %	5.91 %	0.7 px	1.3 px	100.00 %	TBD	NVIDIA TITAN XP	
10	LiteFlowNet3			2.51 %	5.90 %	0.7 px	1.3 px	100.00 %	TBD	NVIDIA TITAN XP	
11	HTC			2.55 %	7.84 %	0.8 px	1.6 px	100.00 %	0.03 s	1 core @ 2.5 Ghz (C/C++)	
12	cvpr-304			2.58 %	5.62 %	0.7 px	1.3 px	100.00 %	-1 s	1 core @ 2.5 Ghz (C/C++)	
13	LiteFlowNet2		code	2.63 %	6.16 %	0.7 px	1.4 px	100.00 %	0.0486 s	GTX 1080 (slower than Pascal Titan X)	

Figure 16. **Screenshot on the KITTI 2012 benchmark (printed as PDF).**

2019/11/23 The KITTI Vision Benchmark Suite

- Download development kit (3 MB)

Our evaluation table ranks all methods according to the number of erroneous pixels. All methods providing less than 100 % density have been interpolated using simple background interpolation as explained in the corresponding header file in the development kit. Legend:

- **D1:** Percentage of stereo disparity outliers in first frame
- **D2:** Percentage of stereo disparity outliers in second frame
- **Fl:** Percentage of optical flow outliers
- **SF:** Percentage of scene flow outliers (=outliers in either D0, D1 or Fl)
- **bg:** Percentage of outliers averaged only over background regions
- **fg:** Percentage of outliers averaged only over foreground regions
- **all:** Percentage of outliers averaged over all ground truth pixels

Note: On 13.03.2017 we have fixed several small errors in the flow (noc+occ) ground truth of the dynamic foreground objects and manually verified all images for correctness by warping them according to the ground truth. As a consequence, all error numbers have decreased slightly. Please download the devkit and the annotations with the improved ground truth for the training set again if you have downloaded the files prior to 13.03.2017 and consider reporting these new number in all future publications. The last leaderboards before these corrections can be found here (optical flow 2015) and here (scene flow 2015). The leaderboards for the KITTI 2015 stereo benchmarks did not change.

Important Policy Update: As more and more non-published work and re-implementations of existing work is submitted to KITTI, we have established a new policy: from now on, only submissions with significant novelty that are leading to a peer-reviewed paper in a conference or journal are allowed. Minor modifications of existing algorithms or student research projects are not allowed. Such work must be evaluated on a split of the training set. To ensure that our policy is adopted, new users must detail their status, describe their work and specify the targeted venue during registration. Furthermore, we will regularly delete all entries that are 6 months old but are still anonymous or do not have a paper associated with them. For conferences, 6 month is enough to determine if a paper has been accepted and to add the bibliography information. For longer review cycles, you need to resubmit your results.

Additional information used by the methods

- Stereo: Method uses left and right (stereo) images
- Multiview: Method uses more than 2 temporally adjacent images
- Motion stereo: Method uses epipolar geometry for computing optical flow
- Additional training data: Use of additional data sources for training (see details)

Evaluation ground truth All pixels **Evaluation area** All pixels

	Method	Setting	Code	Fl-bg	Fl-fg	Fl-all	Density	Runtime	Environment	Compare
1	UberATG-DRISF	Stereo		**3.59 %**	10.40 %	**4.73 %**	100.00 %	0.75 s	CPU+GPU @ 2.5 Ghz (Python)	
W. Ma, S. Wang, R. Hu, Y. Xiong and R. Urtasun: Deep Rigid Instance Scene Flow. CVPR 2019.										
2	DH-SF	Stereo		4.12 %	12.07 %	5.45 %	100.00 %	350 s	1 core @ 2.5 Ghz (Matlab + C/C++)	
3	IAOSF	Stereo		4.56 %	12.00 %	5.79 %	100.00 %	5 min	1 core @ 3.5 Ghz (Matlab + C/C++)	
4	PCF-F			6.05 %	**5.99 %**	6.04 %	100.00 %	0.08 s	GPU @ 2.5 Ghz (Python)	
5	PPAC-HD3			5.78 %	7.48 %	6.06 %	100.00 %	0.14 s	NVIDIA GTX 1080 Ti	
6	MaskFlownet			5.79 %	7.70 %	6.11 %	100.00 %	0.06 s	NVIDIA TITAN Xp	
7	ISF	Stereo		5.40 %	10.29 %	6.22 %	100.00 %	10 min	1 core @ 3 Ghz (C/C++)	
A. Behl, O. Jafari, S. Mustikovela, H. Alhaija, C. Rother and A. Geiger: Bounding Boxes, Segmentations and Object Coordinates: How Important is Recognition for 3D Scene Flow Estimation in Autonomous Driving Scenarios?. International Conference on Computer Vision (ICCV) 2017.										
8	VCN			5.83 %	8.66 %	6.30 %	100.00 %	0.2 s	Titan X Pascal	
G. Yang and D. Ramanan: Volumetric Correspondence Networks for Optical Flow. NeurIPS 2019.										
9	Mono expansion	Stereo		5.83	8.66	6.30	100.00	0.25 s	GPU @ 2.5 Ghz (Python)	

www.cvlibs.net/datasets/kitti/eval_scene_flow.php?benchmark=flow 2/11

Figure 17. **Screenshot on the KITTI 2015 benchmark (printed as PDF).** MaskFlownet-S ranks 14th.