

# Supplementary Material for Unsupervised Learning of Multi-Frame Optical Flow with Occlusions

Joel Janai<sup>1,3</sup> Fatma Güney<sup>4</sup> Anurag Ranjan<sup>2</sup>  
Michael Black<sup>2</sup> Andreas Geiger<sup>1,3</sup>

<sup>1</sup>Autonomous Vision Group, <sup>2</sup>Perceiving Systems Department,

MPI for Intelligent Systems Tübingen

<sup>3</sup>University of Tübingen

<sup>4</sup>Visual Geometry Group, University of Oxford

joel.jana@tue.mpg.de

**Abstract.** The **supplementary video** shows additional qualitative results of our method in comparison with various baselines on the KITTI raw dataset. The **supplementary document** provides details on the parameter settings used throughout our experiments (Section 1) and additional quantitative (Section 2) as well as qualitative (Section 3) results of our approach demonstrating success and failure cases.

## 1 Parameter Settings

For pre-training on RoamingImages and for unsupervised fine-tuning on MPI Sintel [1] and KITTI 2015 [2, 3] we set the hyper-parameters as shown in Table 1. The columns, except for the last two, correspond to the relative weights of different terms in the loss function  $\mathcal{L}(\theta)$  as defined in Eq. (1) in the main paper. In particular, those are the parameters of the photometric loss ( $\omega_P$ ), smoothness constraints ( $\omega_{S1}, \omega_{S2}$ ), the constant velocity constraint ( $\omega_{CV}$ ) and the occlusion prior ( $\omega_O$ ). We use the same parameters for the Clean and Final passes of MPI Sintel.

The column second to the last in Table 1 shows the photometric error function used for the dataset. While the brightness constancy (BC) assumption works well with synthetic data (RoamingImages and MPI Sintel), we utilize the gradient constancy (GC) assumption when training on KITTI since it is more robust to illumination changes which often occur on KITTI.

Finally, we show the order of the smoothness function  $\mathcal{L}_S$  in the last column as mentioned in Section 3.3 of our paper. We use first order smoothness constraints (1st) on RoamingImages and Sintel, and second order smoothness constraints (2nd) on KITTI 2015. The second order smoothness constraint allows piecewise affine flow fields better suited to handle non fronto-parallel surfaces such as the road region in KITTI.

## 2 Quantitative Results

In Table 2, we provide an extended version of the quantitative results table (Table 2 in the main paper) on the MPI Sintel [1] and KITTI 2015 [2, 3] datasets. In addition

**Table 1: Parameter Settings:** In this table, we list the dataset specific hyper-parameters that are used in our experiments: the relative weights of the loss functions in the first five columns, the photometric error function as BC (Brightness Constancy) and GC (Gradient Constancy) in the second to the last column, and the order of the smoothness loss in the last column. Each row corresponds to a dataset.

Dataset	$\omega_P$	$\omega_{S1}$	$\omega_{S2}$	$\omega_{CV}$	$\omega_O$	$\delta(\cdot, \cdot)$	$\mathcal{L}_S$
RoamingImages	2					BC	1st
MPI Sintel	4	0.1	0.1	0.0001	0.1	BC	1st
KITTI 2015	4					GC	2nd

to the results in the paper, Table 2(a) also shows the cross-dataset performance of our approach, i.e. trained on one dataset and tested on another, compared to the previous approaches. Our model fine-tuned on KITTI 2015 performs similarly to the pre-trained model on MPI Sintel and vice versa. This shows the generalization capability of our approach without over-fitting to a specific dataset.

In addition to the overall performance on the test set reported in the paper (Table 2), we report the error in both occluded and non-occluded regions in Table 2(b). After fine-tuning using the soft-constraint, the performance significantly increases in non-occluded regions on all datasets. In occluded regions, there are only minor improvements or even a degradation in performance (Sintel Final). The soft constraint allows deviations from the constant velocity model resulting in improvements in non-occluded regions with complex motion. However, less information is available for occluded regions when switching from the hard-constraint to the soft-constraint. In other words, the predictions rely more on spatial information than on temporal information. Still, the overall performance improves with the soft-constraint since non-occluded regions typically cover a larger area compared to occluded regions.

### 3 Qualitative Results

In this section, we show additional qualitative results of our fine-tuned models on KITTI 2015 (Fig. 1), MPI Sintel Clean (Fig. 2) and Final (Fig. 3). Despite missing explicit supervision, our predictions are mostly accurate. However, large motions and fine details lead to some failure cases (last three rows in Fig. 1, Fig. 2 and Fig. 3).

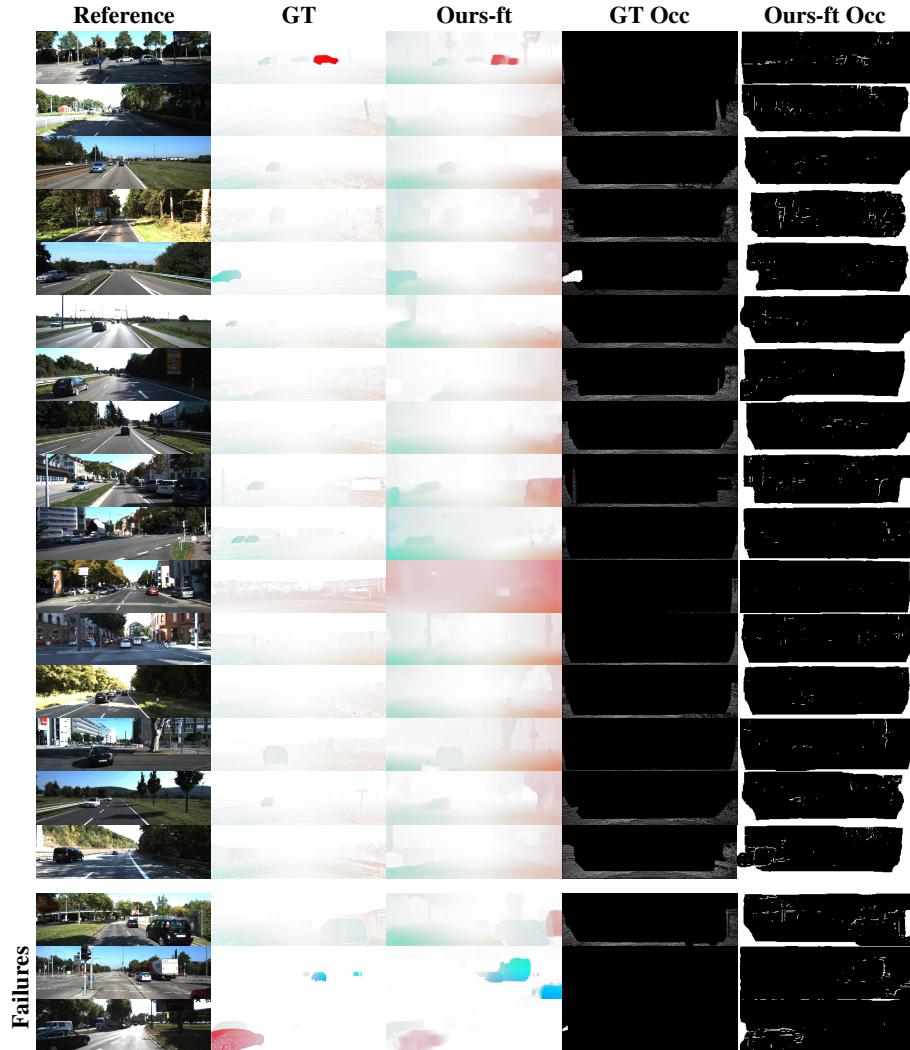
Table 2: **Quantitative Results:** In these tables, we compare our method to several state-of-the-art unsupervised methods on the training sets (a) and test sets (b) of MPI Sintel and KITTI 2015 datasets. We report the Average End-point Error for all pixels (All), non occluded pixels (NOC) and occluded pixels (OCC). For the KITTI 2015 test set (b) we provide the error ratio for all pixels (All) and non-occluded pixels (NOC) since it is the official evaluation measure used by the KITTI benchmark. We use parenthesis to indicate cases where training was performed on the same dataset and mark the cases where only the annotated samples were excluded from training with \*. Missing entries (-) were not reported for the respective method and bold fonts highlight the best results.

	Methods	MPI Sintel Clean			MPI Sintel Final			KITTI 2015		
		All	NOC	OCC	All	NOC	OCC	All	NOC	OCC
Unsupervised	DSTFlow [4]	6.93	5.05	-	7.82	5.97	-	24.30	14.23	-
	DSTFlow-ft-Kitti [4]	7.10	5.26	-	7.95	6.16	-	16.79*	6.96*	-
	DSTFlow-ft-Sintel [4]	(6.16)	(4.61)	-	(6.81)	(4.91)	-	25.98	15.89	-
	UnFlow-CSS [5]	-	-	-	7.91	-	-	8.10*	-	-
	OccAwareFlow [6]	5.23	-	-	6.34	-	-	21.30	-	-
	OccAwareFlow-Kitti-ft [6]	7.41	-	-	7.92	-	-	8.88*	-	-
	OccAwareFlow-Sintel-ft [6]	(4.03)	-	-	(5.95)	-	-	22.6	-	-
	Ours-Hard	5.38	4.32	11.58	6.01	4.92	<b>12.42</b>	15.63	8.80	41.65
	Ours-Soft-Kitti-ft	5.82	4.63	12.89	6.63	5.48	13.60	<b>6.59*</b>	<b>3.22*</b>	<b>19.11*</b>
	Ours-Soft-Sintel-ft	(3.89)	(2.64)	(11.21)	(5.52)	(4.32)	(12.87)	15.69	7.87	46.34

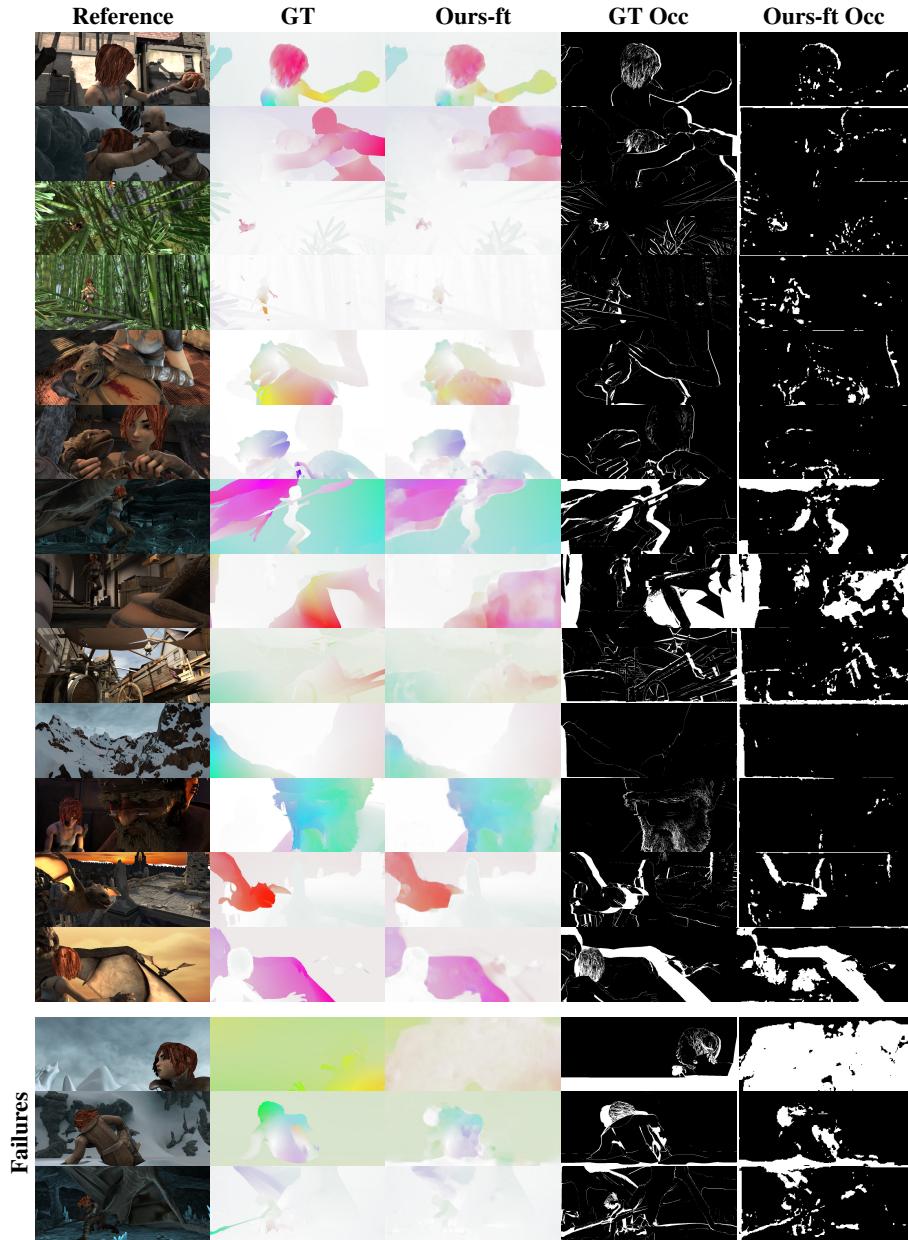
(a) Training

	Methods	MPI Sintel Clean			MPI Sintel Final			KITTI 2015	
		All	NOC	OCC	All	NOC	OCC	All	NOC
Unsupervised	DSTFlow [4]	10.40	5.20	-	11.11	5.92	-	-	-
	DSTFlow-ft-Kitti [4]	10.95	5.87	-	11.80	6.70	-	39.00%	-
	DSTFlow-ft-Sintel [4]	10.41	5.30	-	11.28	6.16	-	-	-
	UnFlow-CSS [5]	9.38	5.37	42.16	10.22	6.06	44.11	23.30%	14.68%
	OccAwareFlow-Kitti-ft [6]	-	-	-	-	-	-	31.20%*	23.53%*
	OccAwareFlow-Sintel-ft [6]	7.95	4.08	39.53	9.15	5.21	41.31	-	-
	Ours-Hard	8.35	4.81	37.14	9.38	5.76	<b>38.84</b>	48.93%	41.09%
	Ours-Soft-Kitti-ft	-	-	-	-	-	-	<b>22.94%</b>	<b>13.85%</b>
	Ours-Soft-Sintel-ft	<b>7.23</b>	<b>3.60</b>	<b>36.78</b>	<b>8.81</b>	<b>5.03</b>	39.65	-	-

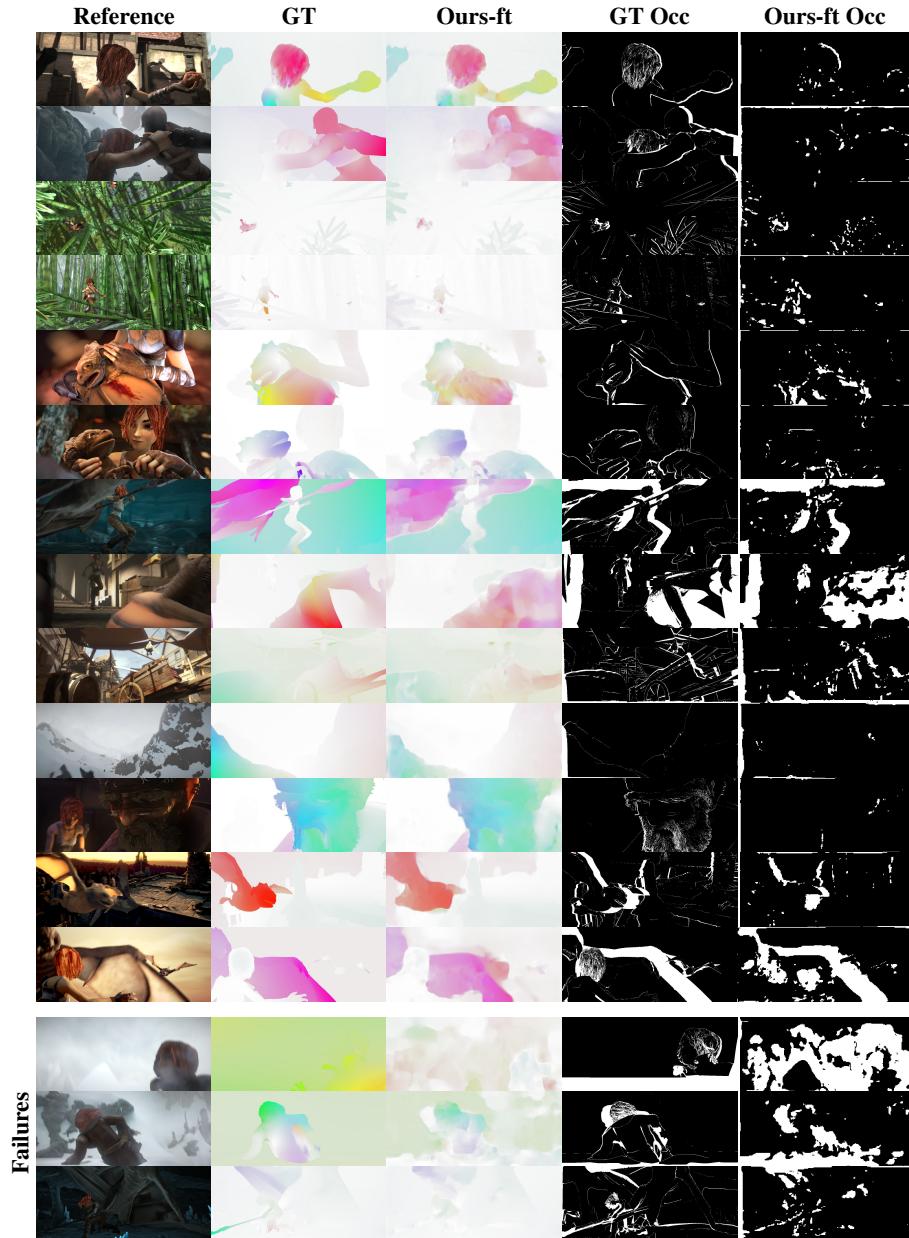
(b) Test



**Fig. 1: Qualitative Results:** In this figure, we show our results with multiple frames and occlusion reasoning (third column) on examples from KITTI 2015. Our model produces accurate flow estimates with sharp boundaries as well as accurate occlusion estimates (last column).



**Fig. 2: Qualitative Results:** In this figure, we show our results with multiple frames and occlusion reasoning (third column) on examples from MPI Sintel Clean. Our model produces accurate flow estimates with sharp boundaries as well as accurate occlusion estimates (last column).



**Fig. 3: Qualitative Results:** In this figure, we show our results with multiple frames and occlusion reasoning (third column) on examples from MPI Sintel Final. Our model produces accurate flow estimates with sharp boundaries as well as accurate occlusion estimates (last column).

## References

1. Butler, D.J., Wulff, J., Stanley, G.B., Black, M.J.: A naturalistic open source movie for optical flow evaluation. In: Proc. of the European Conf. on Computer Vision (ECCV). (2012) 1
2. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: The KITTI dataset. International Journal of Robotics Research (IJRR) **32**(11) (2013) 1231–1237 1
3. Menze, M., Geiger, A.: Object scene flow for autonomous vehicles. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). (2015) 1
4. Ren, Z., Yan, J., Ni, B., Liu, B., Yang, X., Zha, H.: Unsupervised deep learning for optical flow estimation. In: Proc. of the Conf. on Artificial Intelligence (AAAI). (2017)
5. Meister, S., Hur, J., Roth, S.: Unflow: Unsupervised learning of optical flow with a bidirectional census loss. In: Proc. of the Conf. on Artificial Intelligence (AAAI). (2018)
6. Wang, Y., Yang, Y., Yang, Z., Zhao, L., Wang, P., Xu, W.: Occlusion aware unsupervised learning of optical flow. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). (2018)