Imposing Consistency for Optical Flow Estimation
(Supplementary Material)

1. Datasets

In our experiments, we have utilized the FlyingChairs (C) [2], FlyingThings3D (T) [5], Sintel (S) [1], and KITTI (K) [3,6] datasets which are the most popular datasets in the optical flow estimation problem. FlyingChairs [2] consist of 22,872 image pairs and the corresponding ground truths. It is composed of individual pairs, so we cannot constitute additional image pairs corresponding to $k > 1$. FlyingThings3D [5] consists of a training dataset of 21,818 images and a test dataset of 4,248 images. The images of FlyingThings3D consist of more than two consecutive frames, which have both the forward optical flow (20,151 pairs) and the backward optical flow (20,151 pairs) ground truth. In addition, this and Sintel datasets are categorized into clean pass and final pass, and the final pass is applied a post-processing such as fog impact, motion blur, and so on. Therefore, the number of pairs in the training set of FlyingThings3D dataset increases to 80,604. Sintel [1] consists of a training dataset of 1,064 images and a test dataset of 564 images. Sintel is also composed of more than two consecutive frames, and as mentioned above, it is composed of a clean pass and a final pass. KITTI [3,6] consists of a training dataset of 400 images and a test dataset of 400 images. HD1K [4] consists of 1,083 images. These are composed of individual pairs same as FlyingChairs, so there are 200 pairs in both training and test datasets.

Table 1. We perform hyperparameter search over a grid of $\lambda_1 \in \{1.0, 0.1, 0.01, 0.001\}$ in Eq.10. We trained the model with the Flyingchairs (C) and Flyingthings (T) datasets and tested the model on the training dataset of the Sintel (S) and KITTI (T).

Method (small)	λ_2	Sintel (train-EPE)		KITTI-15 (train)	
		Clean	Final	F1-epe	F1-all
RAFT	-	2.21	3.35	7.51	26.9
RAFT + OC	1.0	2.48	3.60	8.57	27.6
	0.1	**2.05**	**3.18**	**7.07**	**23.5**
	0.01	2.19	3.24	7.41	23.6
	0.001	2.24	3.26	7.52	25.0

Table 2. We perform hyperparameter search over a grid of $\lambda_2 \in \{1.0, 0.1, 0.01, 0.001\}$ in Eq.10. The parameters are set to Transformation = R, ϵ = 25.0, and k = 1,2. We trained the model with the Flyingchairs (C) and Flyingthings (T) datasets and tested the model on the training dataset of the Sintel (S) and KITTI (T).

Method (small)	$\lambda 1$	Sintel (train-EPE)		KITTI-15 (train)	
		Clean	Final	F1-epe	F1-all
RAFT	-	2.21	3.35	7.51	26.9
RAFT + TC	1.0	3.05	3.87	13.41	34.7
	0.1	2.06	3.23	7.16	23.3
	0.01	**2.05**	**3.15**	6.50	**22.5**
	0.001	**2.05**	3.20	**6.47**	22.7

Table 3. We perform hyperparameter search over a grid of epsilon $\epsilon \in \{3^2, 5^2, 7^2, \infty\}$ in Eq.8 under Transformation Consistency setting. The parameters in Transformation Consistency are set to λ_2 = 0.01, Transformation = R, and k = 1,2. We trained the model with the Flyingchairs (C) and Flyingthings (T) datasets and tested the model on the training dataset of the Sintel (S) and KITTI (T).

Method (small)	ϵ	Sintel (train-EPE)		KITTI-15 (train)	
		Clean	Final	F1-epe	F1-all
RAFT	-	2.21	3.35	7.51	26.9
RAFT + TC	3^2	2.09	3.19	**6.46**	**22.5**
	5^2	2.05	**3.15**	6.50	**22.5**
	7^2	**2.04**	3.16	6.63	22.6
	∞	2.09	3.18	6.91	22.9

2. Implementation Details

The codes used for our experiments are based on Pytorch, and we have used the official code[1] for RAFT [7]. Our method introduces three additional hyper parameters, namely, (λ_1, λ_2) of Eq.10 and ϵ of Eq.8. We performed a grid search over the values in $\{1.0, 0.1, 0.01, 0.001\}$ for each λ in Eq.10 and in $\{3^2, 5^2, 7^2, \infty\}$ for ϵ value in Eq.8. In table 1, our model with occlusion consistency shows best performance at λ_1 = 0.1. For transformation consistency, our model shows superior scores in most evaluations at λ_2 = 0.01. In case of the ϵ, our transformation consistency loss has shown good performance in Sintel dataset with (5^2 and 7^2 for ϵ) and in KITTI dataset with (3^2 and 5^2 for ϵ). Therefore, we set the parameters to be [(λ_1, λ_2) = (0.1, 0.01), ϵ = 5^2].

[1] https://github.com/princeton-vl/RAFT

3. Dataset Characterization with Displacement Distributions

Fig. 1 below shows cumulative density functions (CDFs) of the ground truth displacements for four popular optical flow datasets. In each plot, we accumulate displacement values symmetrically from -100 to 100 for individual dimensions of (u, v), corresponding to the X and Y axes, excluding larger displacements as outliers. For the FlyingChair dataset, the figure shows that most of the samples are near zero with a relatively small variance. The FlyingThings3D dataset, instead, shows a larger variance than FlyingChair and Sintel. In addition, KITTI appears to have a larger variance than the other datasets, possibly due in part to its smaller frame rates used in this dataset. Another interesting observation from the figures is that, unlike other datasets, KITTI demonstrates significant asymmetry in the flow distribution on the Y axis, as the images are probably dominated by downward movements in the images captured with frontal views of the vehicles.

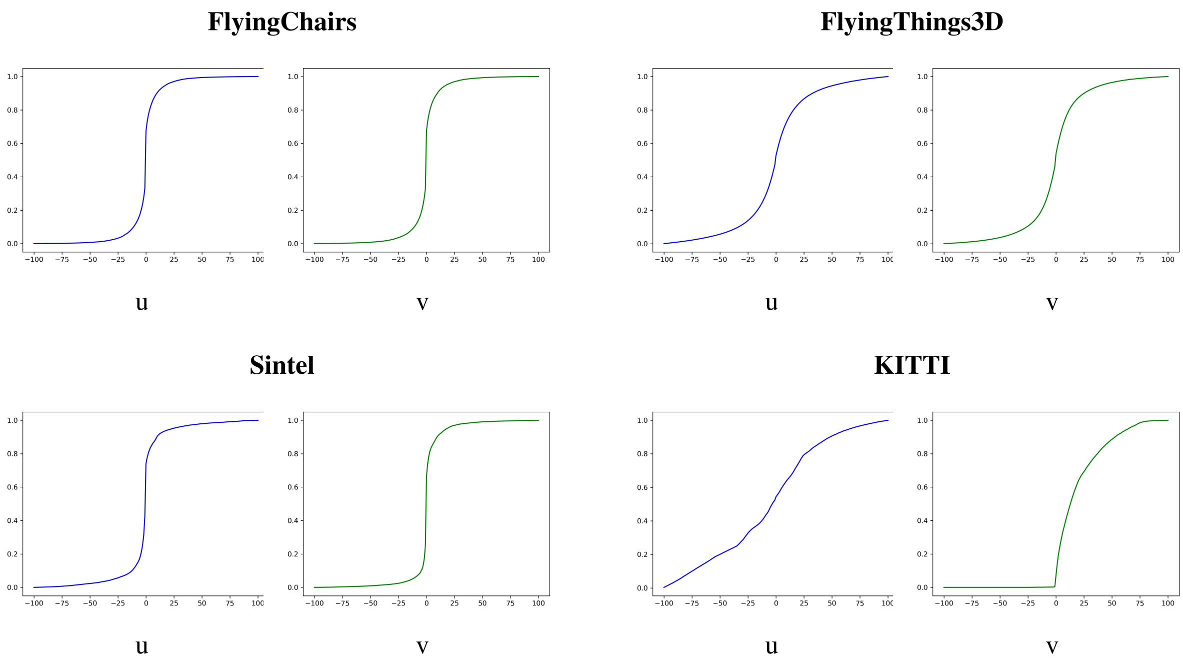

Figure 1. Cumulative Density Functions (CDFs) of displacements in popular datasets. In each pair of sub-figures, the left and rigtht sub-figures show displacements on the X (u) and Y (v) axes, respectively.

4. Performance Comparisons with More Examples

In the figure on the next page, we provide performance comparison with additional examples in a range of various EPEs from low to high.

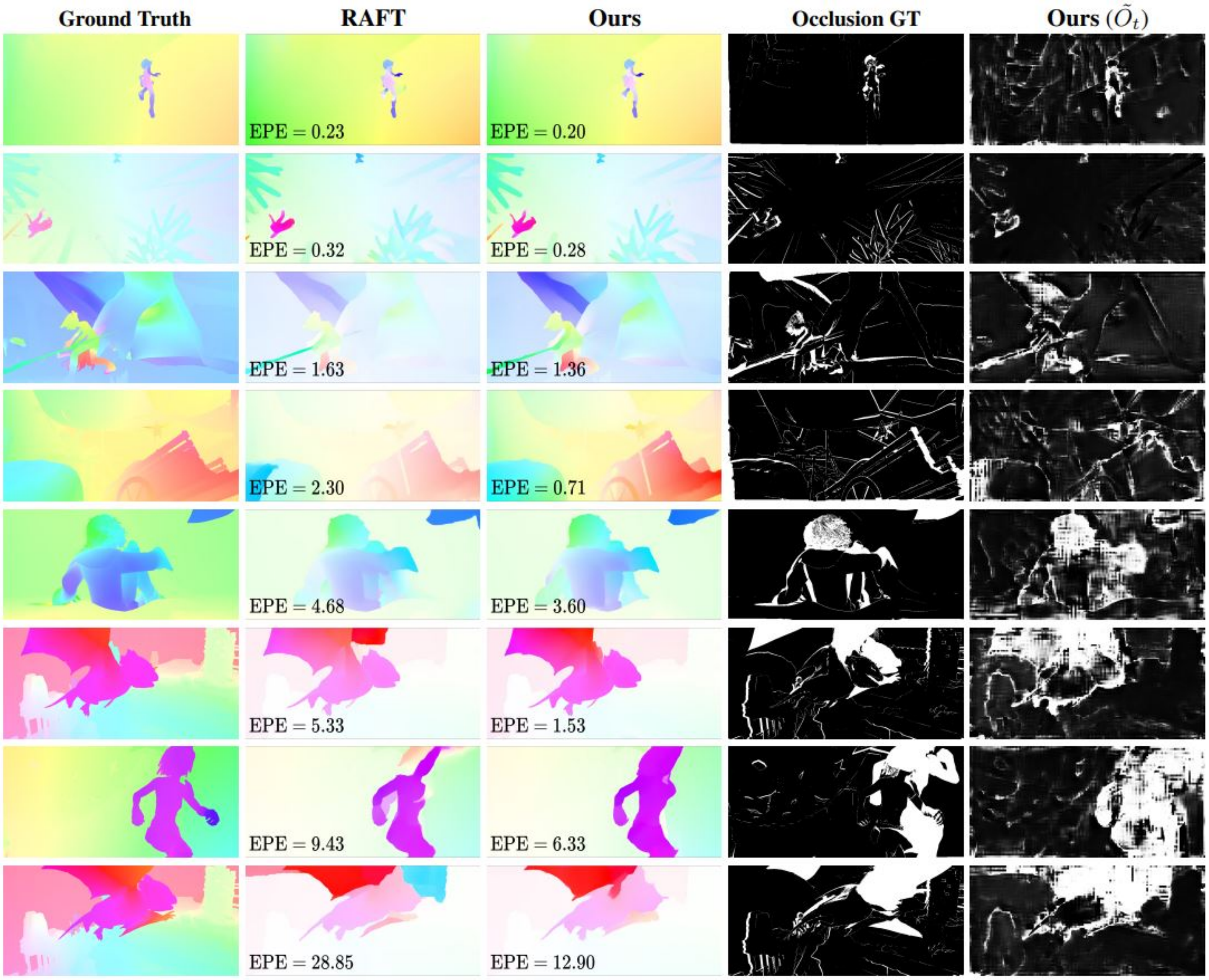

Figure 2. More performance comparisons between the baseline (RAFT) and Ours (RAFT-OCTC) on Sintel train samples (trained with C+T).

References

[1] Daniel J Butler, Jonas Wulff, Garrett B Stanley, and Michael J Black. A naturalistic open source movie for optical flow evaluation. In *European conference on computer vision*, pages 611–625. Springer, 2012. 1

[2] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick Van Der Smagt, Daniel Cremers, and Thomas Brox. Flownet: Learning optical flow with convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2758–2766, 2015. 1

[3] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013. 1

[4] Daniel Kondermann, Rahul Nair, Katrin Honauer, Karsten Krispin, Jonas Andrulis, Alexander Brock, Burkhard Gussefeld, Mohsen Rahimimoghaddam, Sabine Hofmann, Claus Brenner, et al. The hci benchmark suite: Stereo and flow ground truth with uncertainties for urban autonomous driving. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 19–28, 2016. 1

[5] Nikolaus Mayer, Eddy Ilg, Philip Hausser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4040–4048, 2016. 1

[6] Moritz Menze and Andreas Geiger. Object scene flow for autonomous vehicles. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3061–3070, 2015. 1

[7] Zachary Teed and Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow. In *European Conference on Computer Vision*, pages 402–419. Springer, 2020. 1