Complementing Brightness Constancy with Deep Networks for Optical Flow Prediction Supplementary material

Vincent Le Guen[1,2,3], Clément Rambour[2], and Nicolas Thome[2,4]

[1] EDF R&D, Chatou, France
[2] Conservatoire National des Arts et Métiers, CEDRIC, Paris, France
[3] SINCLAIR AI Lab, Palaiseau, France
[4] Sorbonne Université, CNRS, ISIR, F-75005 Paris, France

1 Proof of the uniqueness of the COMBO decomposition

We recall here the COMBO decomposition of the ground truth flow vector $\mathbf{w}^*(\mathbf{x})$ between a physical part $\mathbf{w}_p^*(\mathbf{x})$ fulfilling the Brightness Consistency (BC) assumption, an augmented term $\mathbf{w}_a^*(\mathbf{x})$, and a BC uncertainty term $\alpha^*(\mathbf{x})$:

$$\mathbf{w}^*(\mathbf{x}) = (1 - \alpha^*(\mathbf{x}))\ \mathbf{w}_p^*(\mathbf{x}) + \alpha^*(\mathbf{x})\ \mathbf{w}_a^*(\mathbf{x}). \tag{1}$$

Since the decomposition in Eq. (1) is not necessarily unique, the COMBO decomposition $(\mathbf{w}_p^*, \mathbf{w}_a^*, \alpha^*)$ is defined as the solution of the following constrained optimization problem:

$$\min_{\mathbf{w}_p, \mathbf{w}_a} \quad \|(\mathbf{w}_a, \mathbf{w}_p)\| \qquad \text{subject to:} \tag{2}$$
$$\begin{cases} (1 - \alpha^*(\mathbf{x}))\ \mathbf{w}_p(\mathbf{x}) + \alpha(\mathbf{x})\ \mathbf{w}_a(\mathbf{x}) = \mathbf{w}^*(\mathbf{x}) \\ (1 - \alpha^*(\mathbf{x}))\ |I_1(\mathbf{x}) - I_2(\mathbf{x} + \mathbf{w}_p(\mathbf{x}))| = 0 \\ \alpha^*(\mathbf{x}) = \sigma\left(|I_1(\mathbf{x}) - I_2(\mathbf{x} + \mathbf{w}^*(\mathbf{x}))|\right). \end{cases}$$

We detail here the uniqueness guarantee of the COMBO decomposition in Eq. (2). An unconstrained decomposition would be written as follows:

$$\mathbf{w}(\mathbf{x}) = (1 - \alpha(\mathbf{x}))\ \mathbf{w}_p(\mathbf{x}) + \alpha(\mathbf{x})\ \mathbf{w}_a(\mathbf{x}). \tag{3}$$

It is clear that the naive decomposition in Eq. (3) admits multiple $(\mathbf{w}_p(\mathbf{x}), \mathbf{w}_a(\mathbf{x}), \alpha(\mathbf{x}))$ tuples. We highlight here the effect of the different constraints in Eq. (2) :

- $\alpha^*(\mathbf{x}) = \sigma(\mathcal{L}_{BC}(\mathbf{x}, \mathbf{w}))$ specifies a unique value for $\alpha^*(\mathbf{x})$, but their remains an infinite number of $(\mathbf{w}_p(\mathbf{x}), \mathbf{w}_a(\mathbf{x}))$ tuples.
- $(1 - \alpha(\mathbf{x}))\ |I_1(\mathbf{x}) - I_2(\mathbf{x} + \mathbf{w}_p(\mathbf{x}))| = 0$ specifies a set of BC minimizers for $\mathbf{w}_p$, that we denote $\mathcal{F}_{BC}$.
- By minimizing $||\mathbf{w}_p||$, we limit $\mathbf{w}_p(\mathbf{x}) \in \mathcal{F}_{BC}$ to obtain $\mathbf{w}_p(\mathbf{x}) \in \mathcal{F}_{BC} \cap C_p$, where C_p is the circle of radius $min_{\mathbf{w}_{p'}} ||\mathbf{w}'_p||^2$ (orange in Fig. 1).

– By minimizing $||\mathbf{w}_a||$, we limit $(\mathbf{w}_p(\mathbf{x}), \mathbf{w}_a(\mathbf{x}))$ to the two sets $(\mathbf{w}_p^1(\mathbf{x}), \mathbf{w}_a^1(\mathbf{x}))$ and $(\mathbf{w}_p^2(\mathbf{x}), \mathbf{w}_a^2(\mathbf{x}))$ shown in Fig. 1, which is the intersection between the orange circle and the blue circle of radius $min_{\mathbf{w}_{a'}} ||\mathbf{w}'_a||^2$ in Fig. 1. This constraint following the least action principle, and adds only the minimal information to the BC properly represent $\mathbf{w}(\mathbf{x})$. Finally, by minimizing the angle $\gamma =< \mathbf{w}_p; x >$, we obtain the unique solution $(\mathbf{w}_p^1(\mathbf{x}), \mathbf{w}_a^1(\mathbf{x}))$ shown in Fig 1.

Therefore **the decomposition in Eq. 2 admits a unique tuple** $(\mathbf{w}_p^*, \mathbf{w}_a^*, \alpha^*)$.

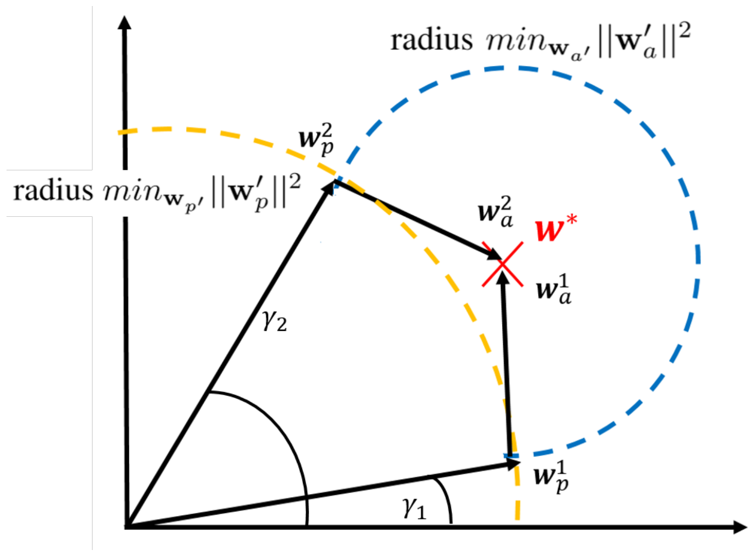

Fig. 1. Illustration of the unique decomposition of COMBO.

2 Implementation details

The code of COMBO is implemented in Pytorch. In the supervised setting, we train models with the following curriculum: FlyingChairs (100000 steps, learning rate=$4e-4$), FlyingThings3D (100000 steps, lr=0.000125), Sintel with additional data from HD1K, FlyingThings and KITTI (100000 steps, lr=0.000125), and finally KITTI (100000 steps, lr=0.001). We use the Adam optimizer with a cyclic learning rate scheduler.

The hyperparameters choosen in the training loss are the following: $\lambda_{total} = 1$, $\lambda_p = 0.1$, $\lambda_a = 0.01$, $\lambda_{photo} = 0.01$, $\lambda_\alpha = 1$, $\lambda_w = 0.1$. This ensures a proper scaling between losses, but we do not attempt at finely tuning the hyperparameters due to the huge computational training time. This setting is kept constant for all stages of the curriculum and datasets.

Each step of the curriculum takes approximately 5 days to converge on a DGX A100 gpu. This gives a very long training curriculum as a whole, highlighting the benefits brought up by the simplification with the semi-supervised training of COMBO.

3 Experiments

3.1 Examples of supervision ($\mathbf{w}_p^*$, $\mathbf{w}_a^*$, α^*)

We provide a few examples of ground truth supervision ($\mathbf{w}_p^*$, $\mathbf{w}_a^*$, α^*) for the datasets FlyingChairs (Figure 2), FlyingThings3D (Figure 3), Sintel (Figure 4) and KITTI (Figure 5).

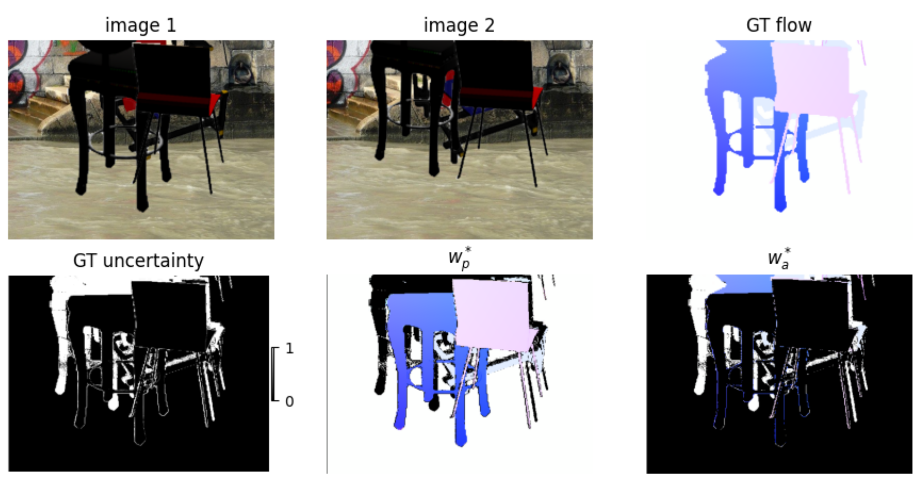

Fig. 2. Example of ground truth supervision ($\mathbf{w}_p^*$, $\mathbf{w}_a^*$, α^*) for FlyingChairs.

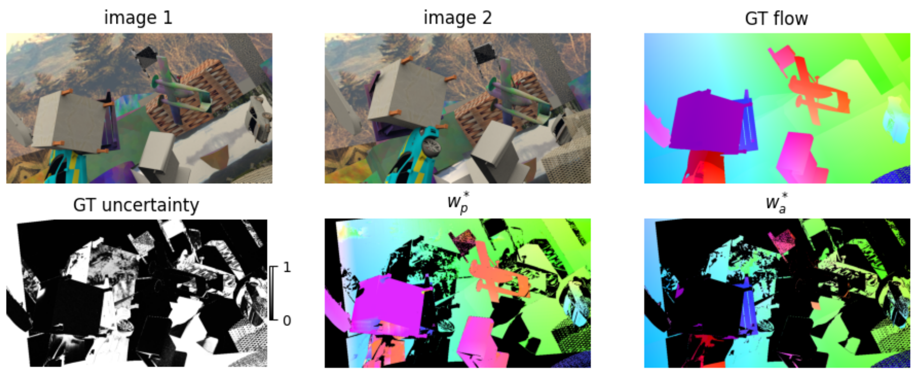

Fig. 3. Example of ground truth supervision ($\mathbf{w}_p^*$, $\mathbf{w}_a^*$, α^*) for FlyingThings3D.

Fig. 4. Example of ground truth supervision ($\mathbf{w}_p^*$, $\mathbf{w}_a^*$, α^*) for Sintel.

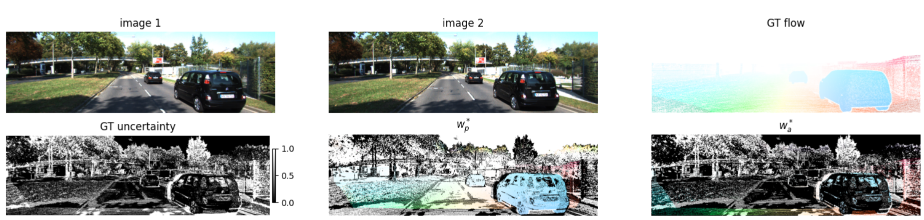

Fig. 5. Example of ground truth supervision ($\mathbf{w}_p^*$, $\mathbf{w}_a^*$, α^*) for KITTI.

3.2 COMBO analysis

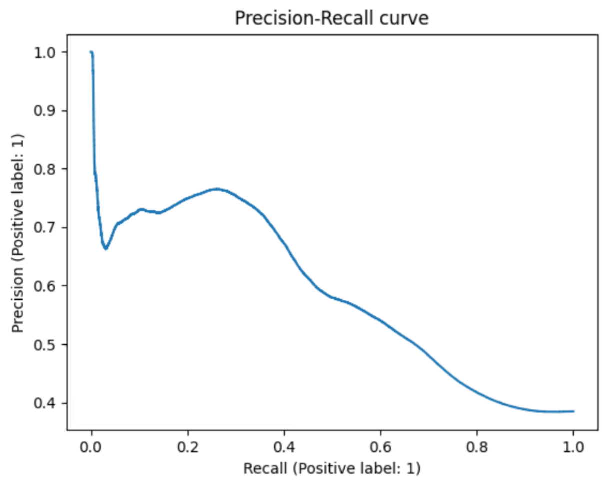

Fig. 6. Precision-recall curve quantifying the ability of COMBO to detect occlusions.

To analyze the ability of COMBO to detect occlusions, we show in Figure 6 the precision-recall curve of the COMBO uncertainty detector with respect to the ground-truth occlusion masks (which ground truth is provided on Sintel). COMBO obtains an Average Precision (AP) of 59% for occlusion detection. This average precision underestimates the true AP since COMBO detects occlusions and also other failure cases: therefore, the precision computed here penalizes a correct BC violation detection (large α) which is not labeled as occlusion. Similarly, some occluded pixels fulfilled the BC assumption, making the recall computed in this manner overestimated. Therefore, the AP reported here is a lower bound of the true AP for occlusions. Despite this, it shows that COMBO is able to efficiently detect occlusions without any ground-truth occlusion supervision, compared to the random classifier which reaches 7% (ratio of occluded pixels).

3.3 Influence of the backbone model

In the main paper, we validate the performances of COMBO based on the RAFT [53] backbone architecture, which is currently one of the state-of-the-art models. However, the COMBO rationale of leveraging the brightness constancy in a deep augmented model is agnostic to the backbone model. We conduct an additional experiment on top of the very recent GMA model [23]. The results shown below (Table 1) on the test set prove that COMBO still provides a significant improvement (epe=0.71 *v.s.* 0.79) compared to this state-of-the-art GMA model on the

FlyingChairs stage, showing that COMBO is a general augmentation strategy for the BC, agnostic to the optical flow architecture.

RAFT	0.82
COMBO (backbone RAFT)	**0.74**
GMA	0.79
COMBO (backbone GMA)	**0.71**

Table 1. Performances of COMBO on the FlyingChairs dataset, based on the RAFT [53] and GMA [23] backbone architectures.

3.4 Additional visualizations

We provide additional visualizations for Sintel in Figures 7, 8, 9 and KITTI-2015 in Figures 10, 11, 12.

In each case, we can observe that in zones of high uncertainty of the brightness constancy, the physical flow $\mathbf{w}_p$ is ill-defined; in these zones, it is efficiently complemented by the flow $\mathbf{w}_a$ to produce an accurate COMBO flow estimate.

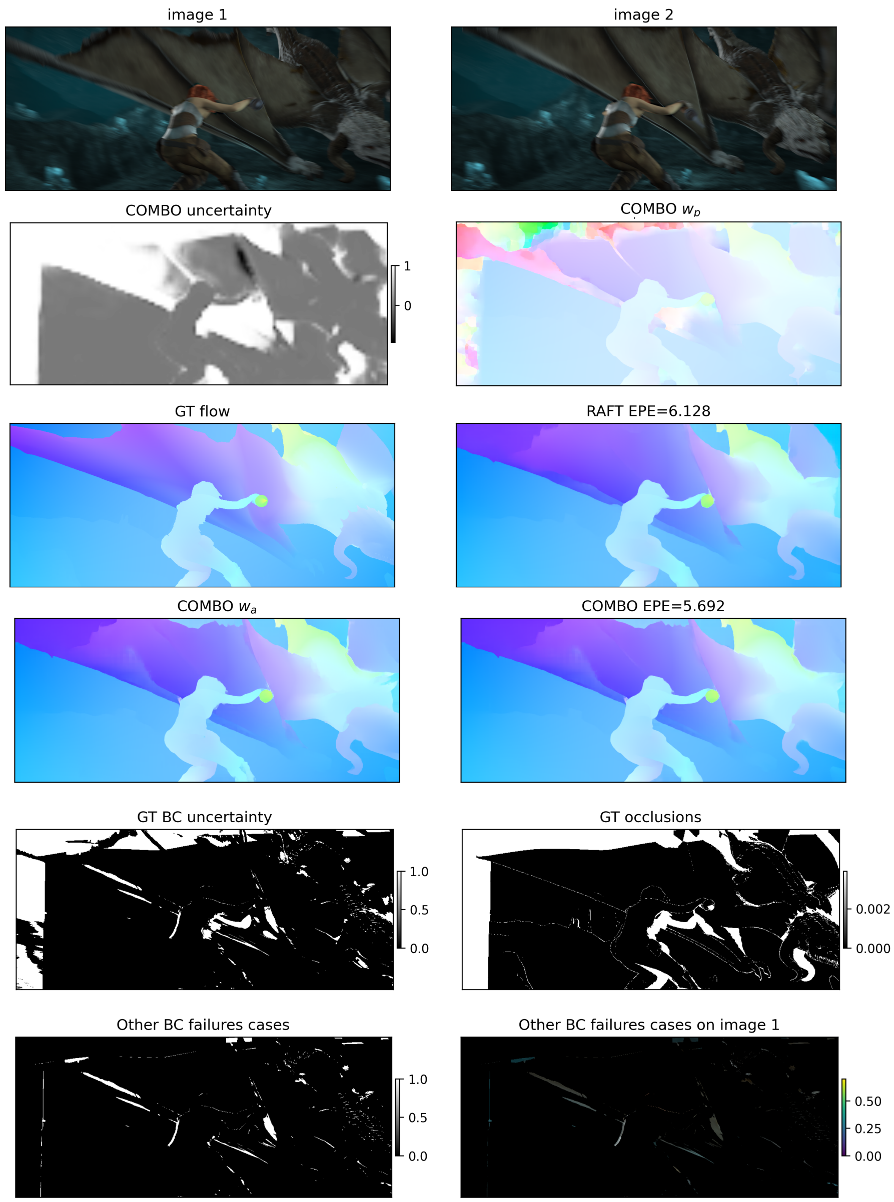

Fig. 7. Additional visualization on Sintel.

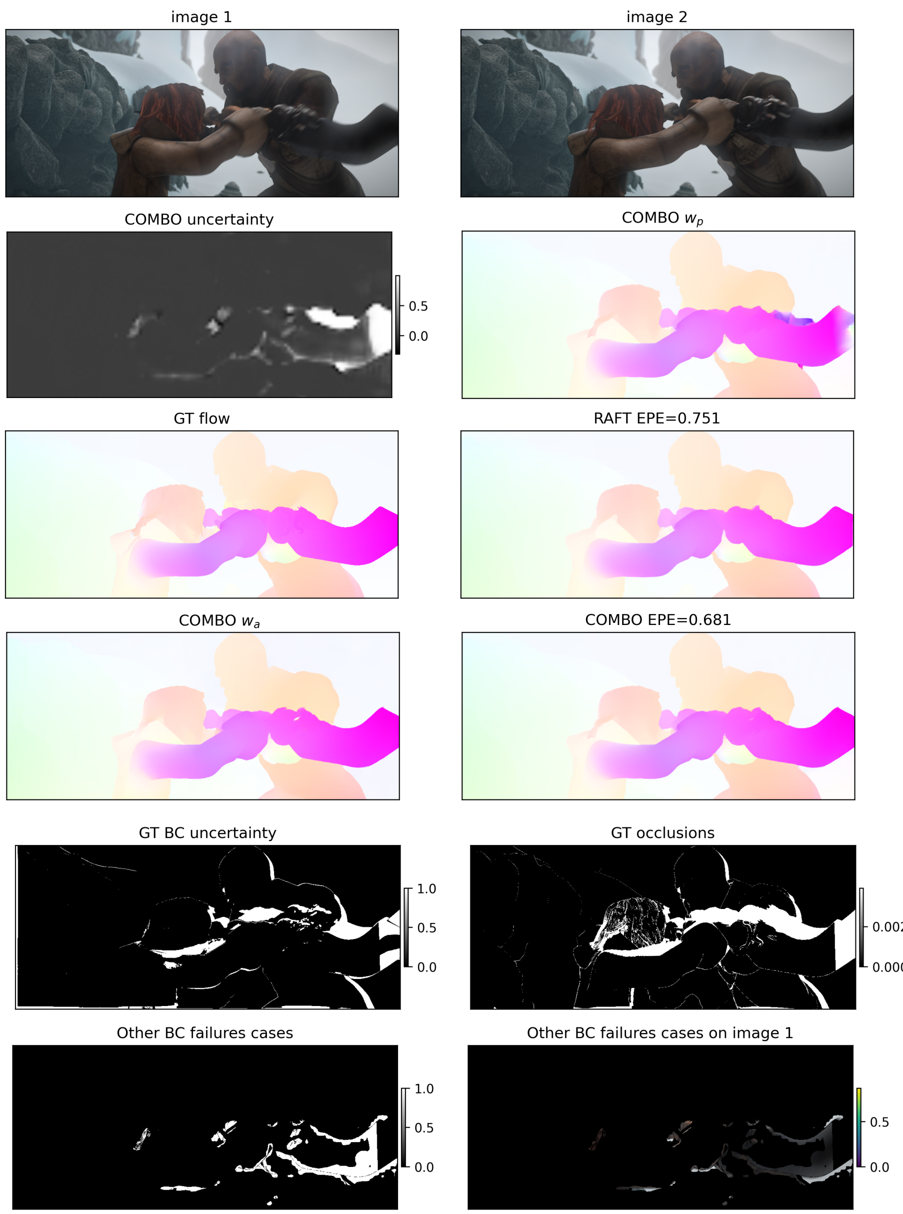

Fig. 8. Additional visualization on Sintel.

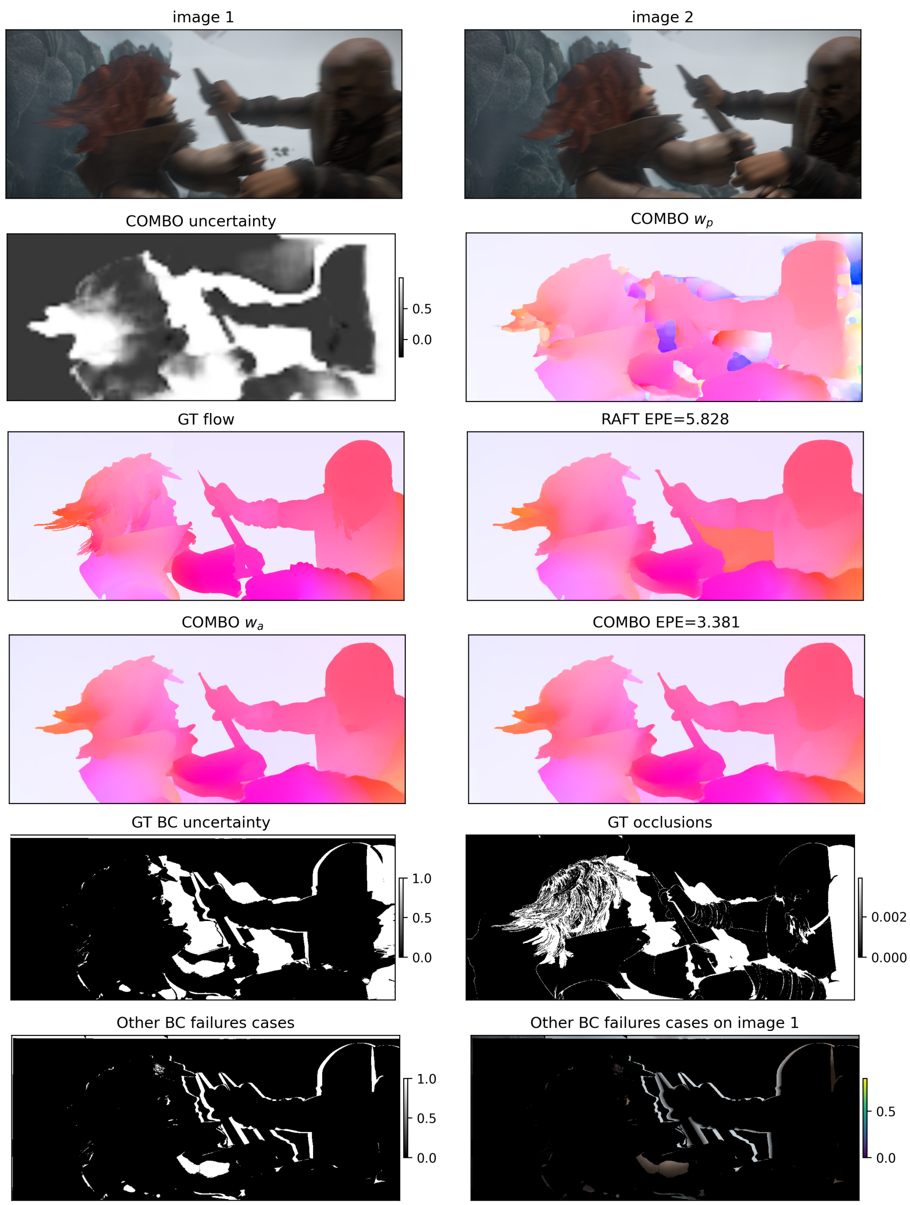

Fig. 9. Additional visualization on Sintel.

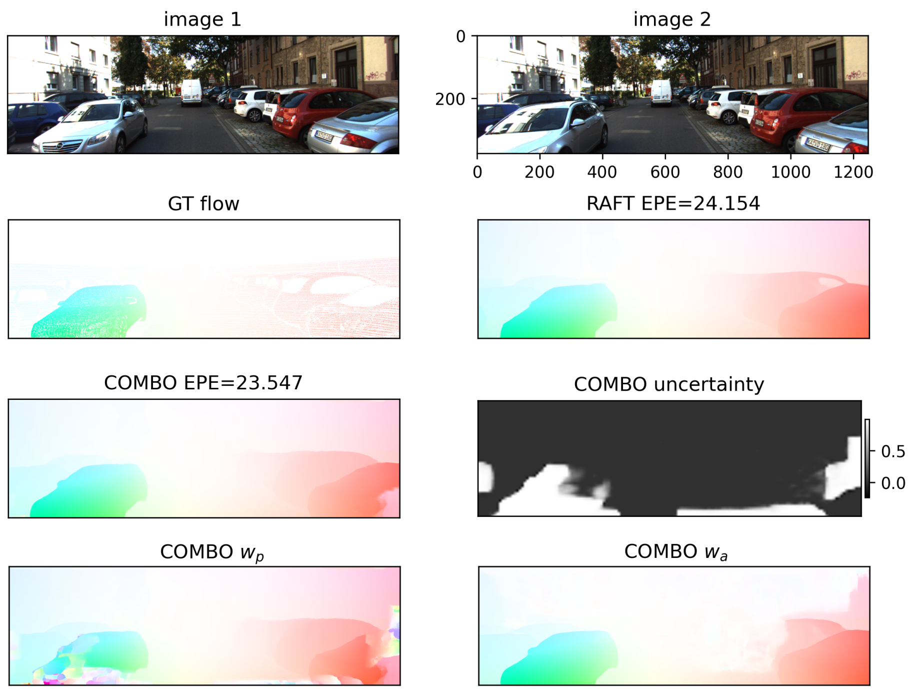

Fig. 10. Additional visualizations on KITTI-2015.

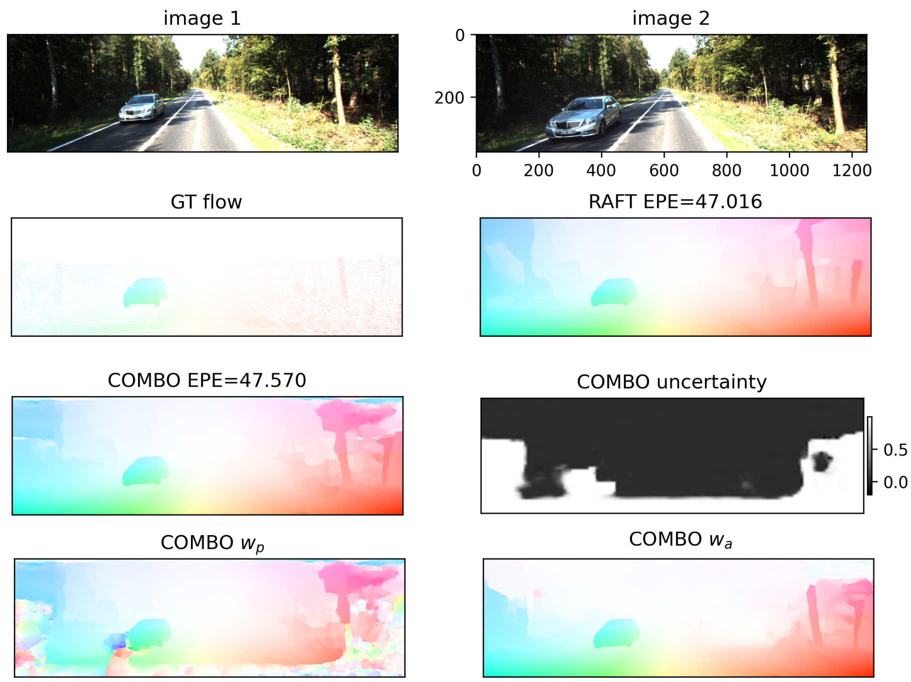

Fig. 11. Additional visualizations on KITTI-2015.

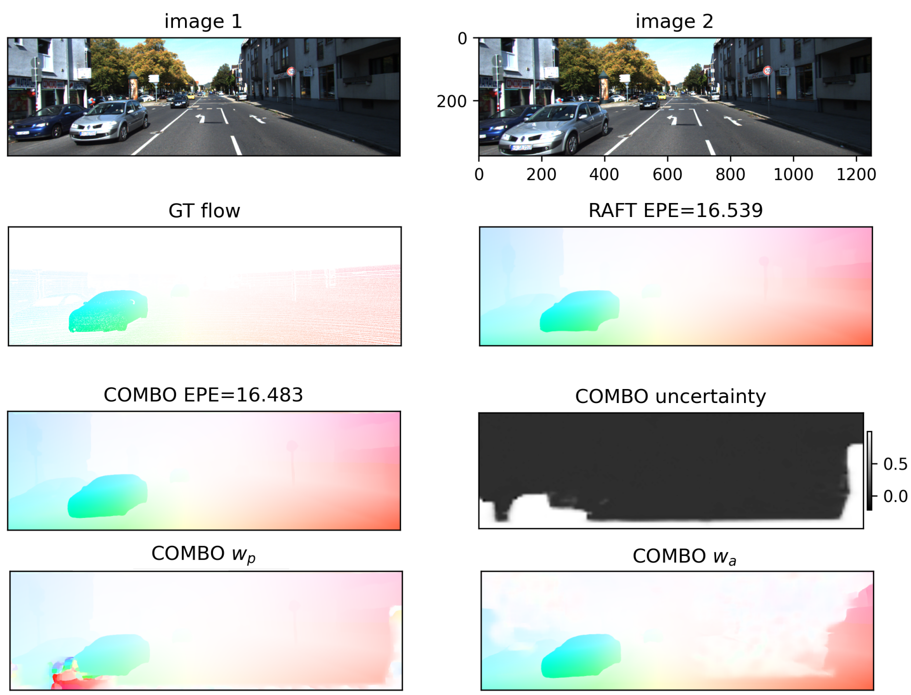

Fig. 12. Additional visualizations on KITTI-2015.

References

1. Agustsson, E., Minnen, D., Johnston, N., Balle, J., Hwang, S.J., Toderici, G.: Scale-space flow for end-to-end optimized video compression. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8503–8512 (2020)
2. Alvarez, L., Deriche, R., Papadopoulo, T., Sánchez, J.: Symmetrical dense optical flow estimation with occlusions detection. International Journal of Computer Vision **75**(3), 371–385 (2007)
3. Bailer, C., Varanasi, K., Stricker, D.: Cnn-based patch matching for optical flow with thresholded hinge embedding loss. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3250–3259 (2017)
4. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: An unsupervised learning model for deformable medical image registration. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 9252–9260 (2018)
5. Bengio, S., Vinyals, O., Jaitly, N., Shazeer, N.: Scheduled sampling for sequence prediction with recurrent neural networks. Advances in neural information processing systems **28** (2015)
6. de Bezenac, E., Pajot, A., Gallinari, P.: Deep learning for physical processes: Incorporating prior scientific knowledge. ICLR (2018)
7. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: European conference on computer vision. pp. 25–36. Springer (2004)
8. Brox, T., Malik, J.: Large displacement optical flow: descriptor matching in variational motion estimation. IEEE transactions on pattern analysis and machine intelligence **33**(3), 500–513 (2010)
9. Bruhn, A., Weickert, J., Schnörr, C.: Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. International journal of computer vision **61**(3), 211–231 (2005)
10. Butler, D.J., Wulff, J., Stanley, G.B., Black, M.J.: A naturalistic open source movie for optical flow evaluation. In: European conference on computer vision. pp. 611–625. Springer (2012)
11. Corpetti, T., Mémin, É., Pérez, P.: Dense estimation of fluid flows. IEEE Transactions on pattern analysis and machine intelligence **24**(3), 365–380 (2002)
12. Daw, A., Thomas, R.Q., Carey, C.C., Read, J.S., Appling, A.P., Karpatne, A.: Physics-guided architecture (pga) of neural networks for quantifying uncertainty in lake temperature modeling. In: Proceedings of the 2020 siam international conference on data mining. pp. 532–540. SIAM (2020)
13. Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., Van Der Smagt, P., Cremers, D., Brox, T.: Flownet: Learning optical flow with convolutional networks. In: Proceedings of the IEEE international conference on computer vision (ICCV). pp. 2758–2766 (2015)
14. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: The kitti dataset. The International Journal of Robotics Research **32**(11), 1231–1237 (2013)
15. Godet, P., Boulch, A., Plyer, A., Le Besnerais, G.: Starflow: A spatiotemporal recurrent cell for lightweight multi-frame optical flow estimation. In: 2020 25th International Conference on Pattern Recognition (ICPR). pp. 2462–2469. IEEE (2021)

16. Greydanus, S., Dzamba, M., Yosinski, J.: Hamiltonian neural networks. In: Advances in Neural Information Processing Systems (NeurIPS). pp. 15353–15363 (2019)
17. Horn, B.K., Schunck, B.G.: Determining optical flow. Artificial intelligence **17**(1-3), 185–203 (1981)
18. Hui, T.W., Tang, X., Loy, C.C.: Liteflownet: A lightweight convolutional neural network for optical flow estimation. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). pp. 8981–8989 (2018)
19. Hur, J., Roth, S.: Iterative residual refinement for joint optical flow and occlusion estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5754–5763 (2019)
20. Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., Brox, T.: Flownet 2.0: Evolution of optical flow estimation with deep networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2462–2470 (2017)
21. Janai, J., Guney, F., Ranjan, A., Black, M., Geiger, A.: Unsupervised learning of multi-frame optical flow with occlusions. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 690–706 (2018)
22. Jason, J.Y., Harley, A.W., Derpanis, K.G.: Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness. In: European Conference on Computer Vision. pp. 3–10. Springer (2016)
23. Jiang, S., Campbell, D., Lu, Y., Li, H., Hartley, R.: Learning to estimate hidden motions with global motion aggregation. CVPR (2021)
24. Jonschkowski, R., Stone, A., Barron, J.T., Gordon, A., Konolige, K., Angelova, A.: What matters in unsupervised optical flow. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16. pp. 557–572. Springer (2020)
25. Kondermann, D., Nair, R., Honauer, K., Krispin, K., Andrulis, J., Brock, A., Gussefeld, B., Rahimimoghaddam, M., Hofmann, S., Brenner, C., 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. pp. 19–28 (2016)
26. Lai, W.S., Huang, J.B., Yang, M.H.: Semi-supervised learning for optical flow with generative adversarial networks. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. pp. 353–363 (2017)
27. Le Guen, V., Thome, N.: Disentangling physical dynamics from unknown factors for unsupervised video prediction. In: CVPR (2020)
28. Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., Anandkumar, A.: Fourier neural operator for parametric partial differential equations. arXiv preprint arXiv:2010.08895 (2020)
29. Linial, O., Eytan, D., Shalit, U.: Generative ODE modeling with known unknowns. ICLR 2020 Deep Differential Equations Workshop (2020)
30. Liu, L., Zhang, J., He, R., Liu, Y., Wang, Y., Tai, Y., Luo, D., Wang, C., Li, J., Huang, F.: Learning by analogy: Reliable supervision from transformations for unsupervised optical flow estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6489–6498 (2020)
31. Liu, P., King, I., Lyu, M.R., Xu, J.: Ddflow: Learning optical flow with unlabeled data distillation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 8770–8777 (2019)

32. Liu, P., Lyu, M., King, I., Xu, J.: Selflow: Self-supervised learning of optical flow. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4571–4580 (2019)
33. Long, Y., She, X., Mukhopadhyay, S.: Hybridnet: integrating model-based and data-driven learning to predict evolution of dynamical systems. Conference on Robot Learning (CoRL) (2018)
34. Lu, L., Jin, P., Karniadakis, G.E.: Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. arXiv preprint arXiv:1910.03193 (2019)
35. Lucas, B.D., Kanade, T., et al.: An iterative image registration technique with an application to stereo vision. Vancouver, British Columbia (1981)
36. Lutter, M., Ritter, C., Peters, J.: Deep lagrangian networks: Using physics as model prior for deep learning. arXiv preprint arXiv:1907.04490 (2019)
37. Mayer, N., Ilg, E., Hausser, P., Fischer, P., Cremers, D., Dosovitskiy, A., Brox, T.: 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. pp. 4040–4048 (2016)
38. Mehta, V., Char, I., Neiswanger, W., Chung, Y., Schneider, J.: Neural dynamical systems. ICLR 2020 Deep Differential Equations Workshop (2020)
39. Meister, S., Hur, J., Roth, S.: Unflow: Unsupervised learning of optical flow with a bidirectional census loss. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)
40. Mémin, E., Pérez, P.: Dense estimation and object-based segmentation of the optical flow with robust techniques. IEEE Transactions on Image Processing **7**(5), 703–719 (1998)
41. Mohan, A.T., Lubbers, N., Livescu, D., Chertkov, M.: Embedding hard physical constraints in neural network coarse-graining of 3d turbulence. arXiv preprint arXiv:2002.00021 (2020)
42. Psichogios, D.C., Ungar, L.H.: A hybrid neural network-first principles approach to process modeling. AIChE Journal **38**(10), 1499–1511 (1992)
43. Raissi, M.: Deep hidden physics models: Deep learning of nonlinear partial differential equations. The Journal of Machine Learning Research **19**(1), 932–955 (2018)
44. Ranjan, A., Black, M.J.: Optical flow estimation using a spatial pyramid network. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). pp. 4161–4170 (2017)
45. Ren, Z., Yan, J., Ni, B., Liu, B., Yang, X., Zha, H.: Unsupervised deep learning for optical flow estimation. In: Thirty-First AAAI Conference on Artificial Intelligence (2017)
46. Rico-Martinez, R., Anderson, J., Kevrekidis, I.: Continuous-time nonlinear signal processing: a neural network based approach for gray box identification. In: Proceedings of IEEE Workshop on Neural Networks for Signal Processing. pp. 596–605. IEEE (1994)
47. Saha, P., Dash, S., Mukhopadhyay, S.: PhICNet: Physics-incorporated convolutional recurrent neural networks for modeling dynamical systems. arXiv preprint arXiv:2004.06243 (2020)
48. Sirignano, J., Spiliopoulos, K.: Dgm: A deep learning algorithm for solving partial differential equations. Journal of computational physics **375**, 1339–1364 (2018)
49. Stone, A., Maurer, D., Ayvaci, A., Angelova, A., Jonschkowski, R.: Smurf: Self-teaching multi-frame unsupervised raft with full-image warping. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3887–3896 (2021)

50. Sun, D., Vlasic, D., Herrmann, C., Jampani, V., Krainin, M., Chang, H., Zabih, R., Freeman, W.T., Liu, C.: Autoflow: Learning a better training set for optical flow. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10093–10102 (2021)
51. Sun, D., Yang, X., Liu, M.Y., Kautz, J.: Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). pp. 8934–8943 (2018)
52. Sun, D., Yang, X., Liu, M.Y., Kautz, J.: Models matter, so does training: An empirical study of cnns for optical flow estimation. IEEE transactions on pattern analysis and machine intelligence **42**(6), 1408–1423 (2019)
53. Teed, Z., Deng, J.: RAFT: recurrent all-pairs field transforms for optical flow. In: European Conference on Computer Vision (ECCV) (2020)
54. Thompson, M.L., Kramer, M.A.: Modeling chemical processes using prior knowledge and neural networks. AIChE Journal **40**(8), 1328–1340 (1994)
55. Wang, Q., Li, F., Tang, Y., Xu, Y.: Integrating model-driven and data-driven methods for power system frequency stability assessment and control. IEEE Transactions on Power Systems **34**(6), 4557–4568 (2019)
56. Wang, S., Wang, H., Perdikaris, P.: Learning the solution operator of parametric partial differential equations with physics-informed deeponets. arXiv preprint arXiv:2103.10974 (2021)
57. Wang, Y., Yang, Y., Yang, Z., Zhao, L., Wang, P., Xu, W.: Occlusion aware unsupervised learning of optical flow. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4884–4893 (2018)
58. Wedel, A., Cremers, D., Pock, T., Bischof, H.: Structure-and motion-adaptive regularization for high accuracy optic flow. In: 2009 IEEE 12th International Conference on Computer Vision. pp. 1663–1668. IEEE (2009)
59. Xu, J., Ranftl, R., Koltun, V.: Accurate optical flow via direct cost volume processing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1289–1297 (2017)
60. Yan, W., Sharma, A., Tan, R.T.: Optical flow in dense foggy scenes using semi-supervised learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13259–13268 (2020)
61. Yang, G., Ramanan, D.: Volumetric correspondence networks for optical flow. Advances in neural information processing systems **32**, 794–805 (2019)
62. Yin, Y., Le Guen, V., Dona, J., Ayed, I., de Bézenac, E., Thome, N., Gallinari, P.: Augmenting physical models with deep networks for complex dynamics forecasting. In: Ninth International Conference on Learning Representations ICLR 2021 (2021)
63. Zhao, S., Sheng, Y., Dong, Y., Chang, E.I., Xu, Y., et al.: Maskflownet: Asymmetric feature matching with learnable occlusion mask. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6278–6287 (2020)