Supplementary Material for LiteFlowNet3: Resolving Correspondence Ambiguity for More Accurate Optical Flow Estimation

Tak-Wai Hui[1] [0000-0002-1441-9289] and Chen Change Loy[2] [0000-0001-5345-1591]

[1] The Chinese University of Hong Kong
[2] Nanyang Technological University
https://github.com/twhui/LiteFlowNet3
twhui@ie.cuhk.edu.hk, ccloy@ntu.edu.sg

1 Color Code for Visualizing Flow Fields

We use the tool provided in Sintel evaluation kit [1] to visualize flow fields in the main paper [2] and supplementary material. Fig. 1 illustrates the color code used in the visualization. Flow direction is encoded with color while magnitude is encoded with color intensity. Particularly, white color at the center corresponds to no visual motion. For the ease of visual evaluation, flow fields among the compared methods are normalized by the maximum flow magnitude of the ground truth (when evaluation is performed on training set) or LiteFlowNet3 [2] (when evaluation is performed on testing set) prior to the computing of flow color.

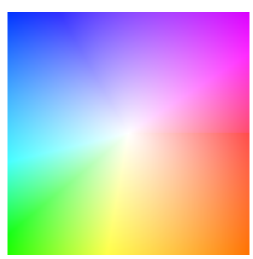


Fig. 1: The color code used in visualizing flow fields.

2 Confidence Predictions

We evaluate the reliability of confidence predictions of LiteFlowNet3 using intersection over union (IoU) on Sintel as HD^3 [11]. Predictions with value smaller than a certain threshold are treated as outliers. For a fair evaluation, we measure IoU on a variant of LiteFlowNet3 [2] that is trained on FlyingChairs without fine-tuning on Sintel. As revealed in Table 1, our confidence prediction achieves

Table 1: Classification results of inlier and outlier predictions in terms of IoU on Sintel training set.

Classes	Sintel Clean	Sintel Final
Outlier	0.259	0.277
Inlier	0.939	0.916
Mean	0.599	0.597

Table 2: AEE results of LiteFlowNet3 trained on FlyingChairs using different training protocols.

Method	Sintel		KITTI 2012	KITTI 2015
	Clean	Final		
Training from scratch	2.59	3.94	3.90	10.51
Using pre-trained model [2]	2.59	3.91	3.88	10.40

reasonably good IoU results. The current CNN implementation uses a single convolution layer for confidence prediction. The classification accuracy can be further improved if a more specialized design is used. It should be noted that our evaluation is not directly comparable to HD^3 as the confidence output of LiteFlowNet3 is trained on a custom-made ground truth. Besides, LiteFlowNet3 is only trained on FlyingChairs for the confidence evaluation while HD^3 is fine-tuned on more datasets.

Examples of optical flow and confidence predictions that are generated by different variants of LiteFlowNet3 are shown in Fig. 2. The flow fields, as illustrated in Figs. 2c and 2d, are generated by the variants trained on FlyingChairs. While Fig. 2e illustrates the flow fields resulting from the variant with fine-tuning on Sintel. The confidence prediction, as shown in Fig. 2g, pinpoints the locations having low confidence (represented by dark color) in the corresponding flow field in Fig. 2c. After incorporating confidence map, the flow field (Fig. 2d) has not only visual appearance closer to the ground truth (Fig. 2f) but also a lower average end-point error (AEE) than the counterpart not using confidence prediction (Fig. 2c).

The above evaluation suggests that the use of confidence prediction is important for LiteFlowNet3 to achieve high optical flow accuracy.

3 Another Training Protocol

Besides plugging-in new modules into a pre-trained LiteFlowNet2 [4] for a faster training, we also study the training of LiteFlowNet3 [2] from scratch on FlyingChairs using the stage-wise training protocol as LiteFlowNet2. Table 2 reveals that both the training protocols result in similar performance.


(a) First frame

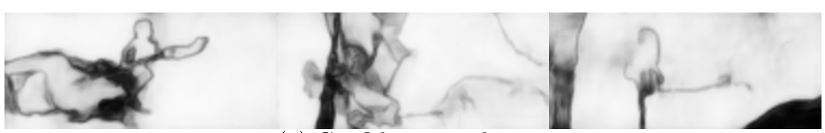

(b) Second frame

(c) No confidence map

(d) With confidence map

(e) With fine-tuning


(f) Flow ground truth


(g) Confidence prediction


(h) Confidence label

Fig. 2: Examples of optical flow and confidence predictions on the training set of Sintel Final.

Table 3: AEE results of optical flow CNNs with our proposed modules trained on FlyingChairs.

Method	Sintel		KITTI 2012	KITTI 2015
	Clean	Final		
FlowNetC [6]	3.77	5.21	6.96	13.53
FlowNetC [6] (with our modules)	3.58	5.04	6.52	12.96
PWC-Net [9]	3.12	4.37	4.53	12.44
PWC-Net [9] (with our modules)	3.05	4.26	4.37	12.28

4 Generalization to Other Optical Flow CNNs

We evaluate the effectiveness of the proposed modules on FlowNetC [6] and PWC-Net [9]. Table 3 suggests that our modules lead to improvement in other architectures as well.

5 Additional Visual Results

Here, we further compare the visual performance of LiteFlowNet3 against other state-of-the-art methods. These include SPyNet [8], FlowNet2 [6], PWC-Net[3] [9], PWC-Net+ [10], IRR-PWC-Net [5], HD^3 [11], SelFlow [7], LiteFlowNet [3], LiteFlowNet2 [4], and LiteFlowNet3 [2]. Except SelFlow, all the above CNNs are 2-frame methods.

Sintel. As shown in Fig. 3, LiteFlowNet3 not only outperforms the compared methods in terms of AEE but also produces superior flow fields. For the first example in Fig. 4, only FlowNet2 and LiteFlowNet3 correctly infer the visual motion of the man standing in the center. For the second example in the same figure, the visual motion of the girl is more accurately recovered by LiteFlowNet3 than the other compared methods. However, artifacts are observed near the legs and head in the flow field resulting from HD^3.

KITTI. As shown in Fig. 5, LiteFlowNet3 achieves the best AEE among the compared methods. Particularly, the second example in Fig. 5, only LiteFlowNet3 correctly generates the visual motion of the traffic pole and nearby region. However, the flow field resulting from HD^3 contains artifacts in the sky. For the first example in Fig. 6, both PWC-Net+, HD^3, and LiteFlowNet3 correctly estimate the visual motion of the traffic pole and nearby region. However, flow bleeding or blurring near the traffic pole is observed in the flow fields resulting from the other methods. For the second example in the same figure, the visual motions of the car and nearby region are more accurately recovered by PWC-Net+, IRR-PWC, HD^3, and LiteFlowNet3 than the other compared methods. However, IRR-PWC cannot correctly infer the background flow near the tree. Serious artifact is observed in the sky of the flow field resulting from HD^3.

[3] The pre-trained model (available at https://github.com/NVlabs/PWC-Net) uses a larger feature encoder. The number of model parameters increases to 9.37M.

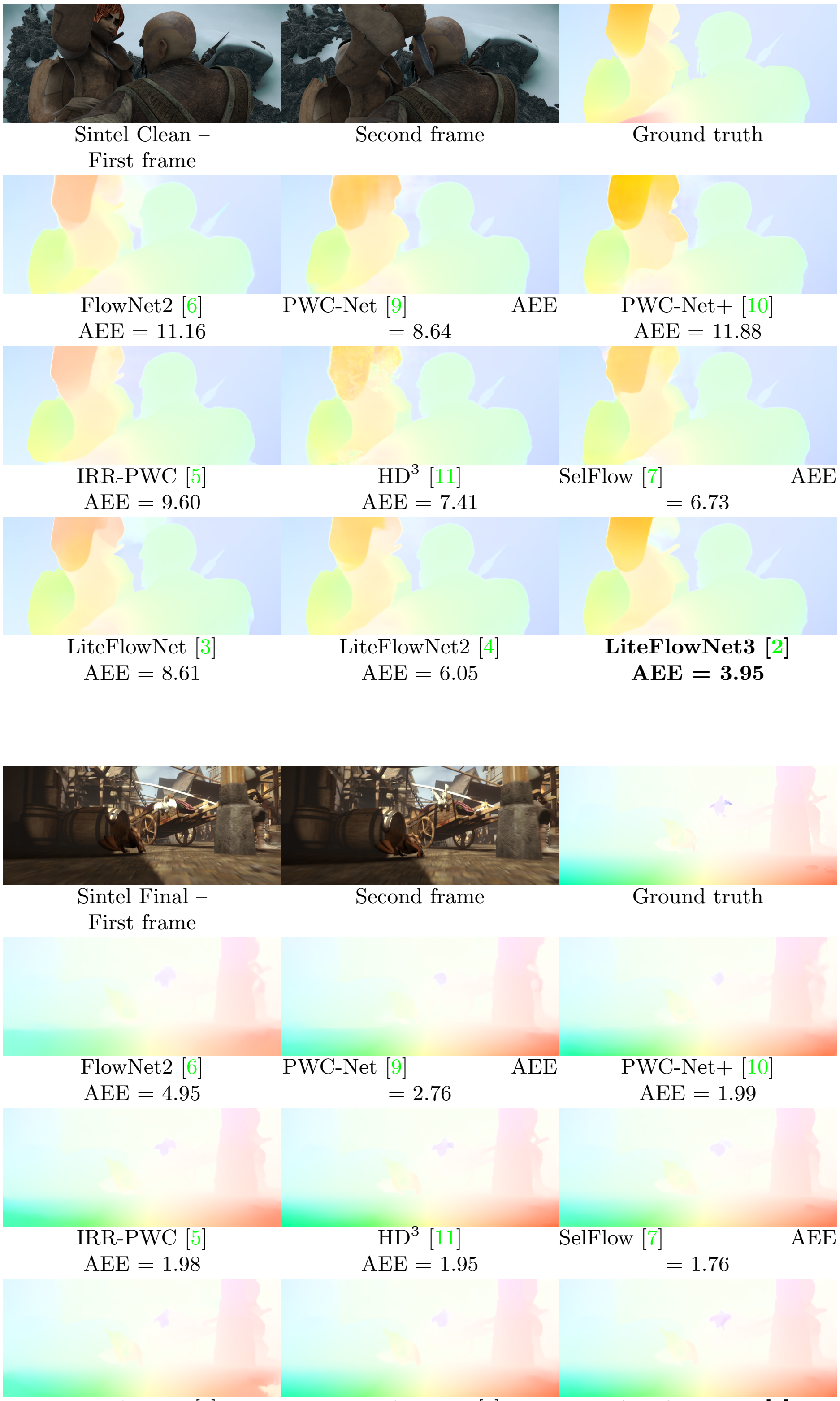


Fig. 3: Examples of optical flow predictions on the training set of Sintel.

Fig. 4: Examples of optical flow predictions on the testing set of Sintel.

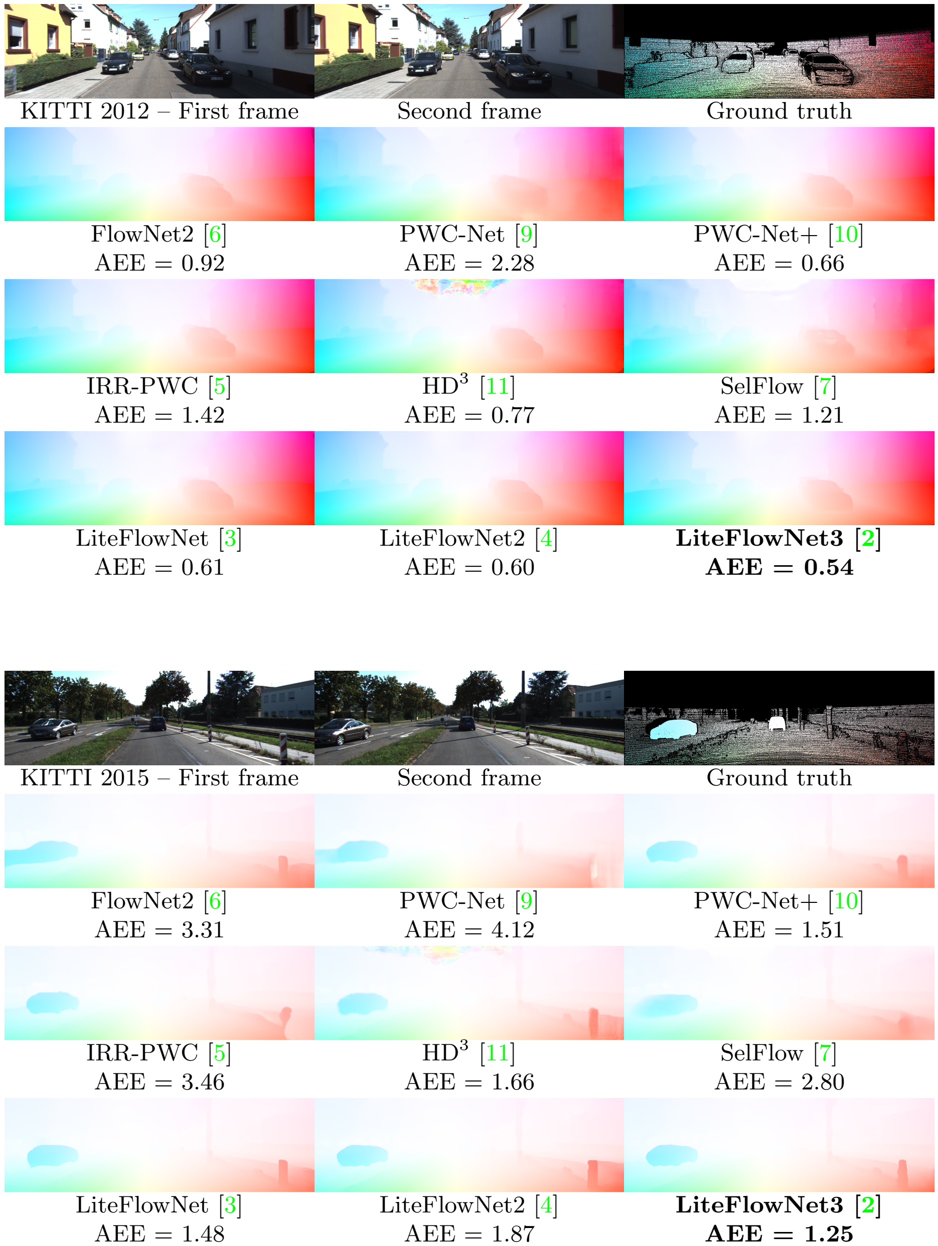


Fig. 5: Examples of optical flow predictions on the training set of KITTI.

Fig. 6: Examples of optical flow predictions on the testing set of KITTI.

References

1. Butler, D.J., Wulff, J., Stanley, G.B., Black, M.J.: A naturalistic open source movie for optical flow evaluation. ECCV pp. 611–625 (2012)
2. Hui, T.W., Loy, C.C.: LiteFlowNet3: Resolving correspondence ambiguity for more accurate optical flow estimation. ECCV (2020)
3. Hui, T.W., Tang, X., Loy, C.C.: LiteFlowNet: A lightweight convolutional neural network for optical flow estimation. CVPR pp. 8981–8989 (2018)
4. Hui, T.W., Tang, X., Loy, C.C.: A lightweight optical flow CNN – Revisiting data fidelity and regularization. TPAMI (2020). https://doi.org/10.1109/TPAMI.2020.2976928
5. Hur, J., Roth, S.: Iterative residual refinement for joint optical flow and occlusion estimation. CVPR pp. 5754–5763 (2019)
6. Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., Brox, T.: FlowNet2.0: Evolution of optical flow estimation with deep networks. CVPR pp. 2462–2470 (2017)
7. Liu, P., Lyu, M., King, I., Xu, J.: SelFlow: Self-supervised learning of optical flow. CVPR pp. 4566–4575 (2019)
8. Ranjan, A., Black, M.J.: Optical flow estimation using a spatial pyramid network. CVPR pp. 4161–4170 (2017)
9. Sun, D., Yang, X., Liu, M.Y., Kautz, J.: PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume. CVPR pp. 8934–8943 (2018)
10. Sun, D., Yang, X., Liu, M.Y., Kautz, J.: Models matter, so does training: An empirical study of CNNs for optical flow estimation. TPAMI (2019). https://doi.org/10.1109/TPAMI.2019.2894353
11. Yin, Z., Darrell, T., Yu, F.: Hierarchical discrete distribution decomposition for match density estimation. CVPR pp. 6044–6053 (2019)