Unsupervised Cumulative Domain Adaptation for Foggy Scene Optical Flow
Supplementary Material

Hanyu Zhou[1], Yi Chang[1]*, Wending Yan[2], Luxin Yan[1]
[1] National Key Laboratory of Science and Technology on Multispectral Information Processing, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology
[2] Huawei International Co. Ltd.
{hyzhou, yichang, yanluxin}@hust.edu.cn, yan.wending@huawei.com

1. Overview

In this supplementary, we first show the effectiveness of cumulative domain adaptation architecture. We also study the effectiveness of depth association module and spatial context attention module. We compare the inference time of different methods. Moreover, we conduct ablation studies on the main parameter setup of the correlation distribution alignment module and the weight sensitivity of model losses. And then, we provide the qualitative comparisons of the proposed UCDA-Flow and other state-of-the-art methods on the synthetic dataset and real foggy images.

2. Discussion

2.1. Effectiveness of Cumulative Adaptation

In Fig. 1, we study the effectiveness of the proposed cumulative depth-association motion adaptation (DAMA) and correlation-alignment motion adaptation (CAMA) for optical flow under real foggy scenes. Without motion adaptation in Fig. 1 (b), there are obvious artifacts in the motion boundaries. With DAMA only in Fig. 1 (c), most of the outliers caused by fog degradation are removed, but there exist some mismatched optical flow values. With both DAMA and CAMA in Fig. 1 (d), the motion results are global-smooth but boundary-sharp. Therefore, the proposed cumulative domain adaptation architecture could benefit us to obtain the final satisfactory result under real foggy scenes.

2.2. Effectiveness of Depth Association

Depth has two constraints for our framework. For the first one, depth is constrained by the principle of the photometric assumption and the smoothness assumption in the stereo match. For the second, there is a natural geometry projection relationship between depth and optical flow, which is used to enhance the motion boundaries in rigid regions and constrain depth estimation. In Fig. 2, we illustrate the effectiveness of depth association. Without geometry depth association, optical flow boundaries of clean images are blurry. With geometry depth association, the motion boundaries are clearer.

Table 1. The inference time on image size 800 × 600.

Method	Paradim	Time (s)	EPE
RobustFlow	Optimization	118	12.25
DenseFogFlow	Semi-Supervised	0.074	1.78
GyroFlow	Weakly-Supervised	**0.055**	0.95
UCDA-Flow	Unsupervised	0.067	**0.81**

2.3. Effectiveness of Spatial Context Attention

In Fig. 3, we simplify the spatial context attention (**SCA**) module in the form of a schematic diagram. Given two adjacent frames, existing optical flow methods directly compute the temporal cost volume cv_{temp} of the two warped frames, and then estimate the optical flow. As shown in Fig. 3 (d), the optical flow has local outliers due to the local mismatches between adjacent frames with the temporal cost volume only.

However, we observe that there are non-local similar objects in the same image. For example, in Fig. 3 (a), the window at the front of the bus has similar contexts with windows in other regions. This observation motivates us to propose the **SCA** module. Specifically, we first sample similar objects by the KNN algorithm, and compute their spatial cost volume, namely cv_{spa}. And then, we take the spatial cost volume cv_{spa} as the correlation saliency attention, and fuse it into the temporal cost volume cv_{temp} to refine the final cost volume $\hat{cv}$ for optical flow in Fig. 3 (c). The proposed spatial context attention module is beneficial to reduce optical flow anomalies caused by mismatch.

2.4. Inference Time

In Table 1, on image size 800 × 600, we compare the inference time of our proposed UCDA-Flow with that of other state-of-the-art methods, including semi-supervised

*Corresponding author.

DenseFogFlow, optimization-based RobustFlow, weakly-supervised GyroFlow. The optimization-based method RobustFlow is time-consuming. The leaning-based methods can reduce inference time. GyroFlow has the shortest inference time compared with DenseFogFlow and the proposed UCDA-Flow. The reason is that GyroFlow can directly obtain the background motion of optical flow from gyroscope data, significantly saving the inference time. On the contrary, although the inference speed of the proposed UCDA-Flow is not the fastest, the performance is the best.

2.5. Parameters of Correlation Distribution Aligned

To study the optimal values for sample number *N* and classes *k* of correlation distribution alignment (**CDA**) module, we select [3, 5, 10, 13] and [100, 1000, 2000, 4000] as candidate values for sample number *N* and classes *k* of **CDA** in Fig. 4. We use KL divergence as the distribution distance metric. In Fig. 4 (a), the original correlation distribution distance between synthetic and real foggy domain is 0.145. In Fig. 4 (b) - (e), we can observe that a large class parameter *k* is conducive to correlation distributions of both the domains close. When *k* is 13, the distributions of both the domains in the high correlation range are close, but in the low correlation range enlarge. Therefore, we choose 10 as the setup of the classes *k*. In Fig. 4 (f) - (i), the larger the sample number *N*, the lower the KL divergence, and the closer the correlation distributions of synthetic and real foggy domains. Since the larger sample number will increase the cost time of model training, we make a trade-off between effectiveness and efficiency. We thus set the sample number *N* as 1000.

2.6. Weight Sensitivity of Model Losses

To choose the optimal weight parameters for the total loss, we conduct the ablation study on the weight sensitivity of the typical model losses, such as $\mathcal{L}_{flow}^{geo}$, $\mathcal{L}_{flow}^{consis}$, $\mathcal{L}_{flow}^{self}$, $\mathcal{L}_{corr}^{kl}$. $\mathcal{L}_{flow}^{geo}$ is to enhance optical flow in rigid regions. $\mathcal{L}_{flow}^{consis}$ is to transfer motion knowledge from clean domain to synthetic foggy domain. The intention of $\mathcal{L}_{flow}^{self}$ and $\mathcal{L}_{corr}^{kl}$ is to transfer motion knowledge from synthetic foggy domain to real foggy domain. In Fig. 5 (a), the larger the weight of $\mathcal{L}_{flow}^{geo}$, the more the depth dominates the network training, resulting in that optical flow may focus too much on rigid regions, but ignore the foreground non-rigid moving objects. In Fig. 5 (b), when the weight of the flow consistency loss $\mathcal{L}_{flow}^{consis}$ is greater than 1, the convergence speed of optical flow network reaches the bottleneck. In Fig. 5 (c), the weight of $\mathcal{L}_{flow}^{self}$ can speed up the knowledge transfer from synthetic foggy domain to real foggy domain. When the weight is greater than 1, the convergence speed of optical flow network training is unchanged. In Fig. 5 (d), we can observe that the correlation distribution alignment loss $\mathcal{L}_{corr}^{kl}$ is sensitive to the framework training. If the weight is too large, the gradient will disappear. Therefore, we set the adaptation losses weights as $\{\lambda_3, \lambda_4, \lambda_5, \lambda_6\} = \{0.1, 1, 1, 0.1\}$.

3. Comparison Experiments

3.1. Qualitative Comparisons on Synthetic Images

The visual results of optical flow predicted by the UCDA-Flow and other state-of-the-art approaches on the synthetic Fog-KITTI2015 dataset are shown in Fig. 6. We choose RobustFlow, UFlow, and Selflow for fair comparison which all do not need any ground truth of optical flow. We take AECR-Net as the defog method to make the state-of-the-art methods trained on the defogging results of Fog-KITTI2015. We can observe that RobustFlow can not work normally. The unsupervised method UFlow contains degradation, especially for motion boundary. Although the Defog + UFlow/Selflow can obtain sharp boundaries, there are artifacts caused by the residual degradation. On the contrary, our results are visual pleasure with global smoothness and sharp boundaries.

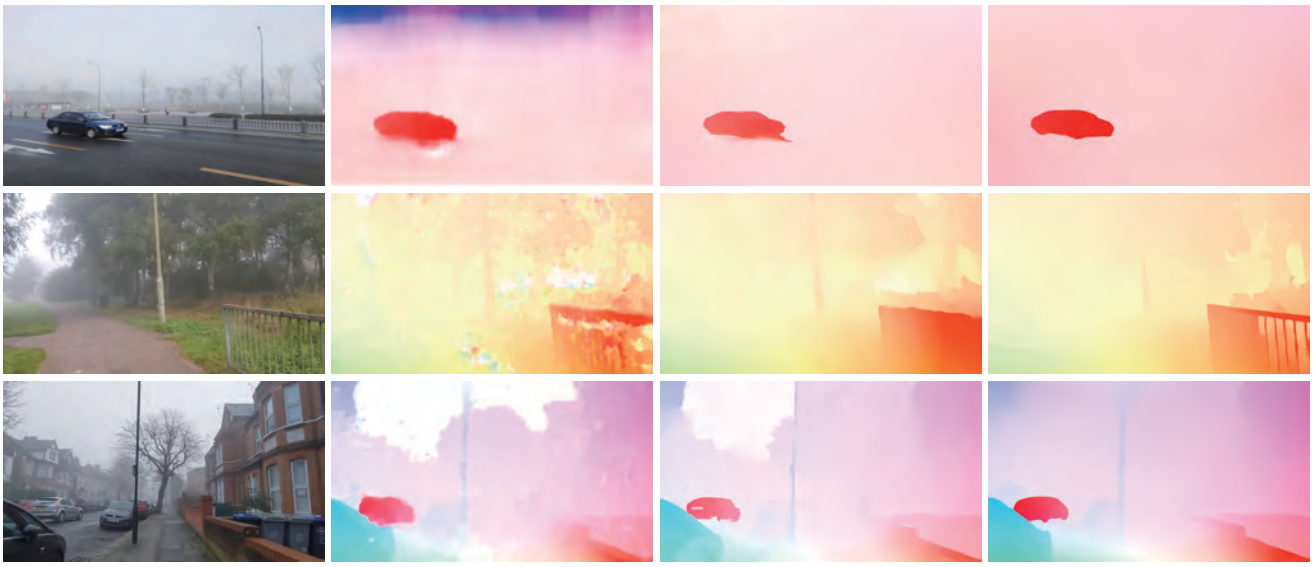
3.2. Qualitative Comparisons on Fog-GOF Dataset

We also provide the visual results of optical flow predicted by different optical flow approaches on the real Fog-GOF dataset in Fig. 7. We choose RobustFlow, DenseFogFlow, and GyroFlow for comparisons, which are designed for adverse weather scenes. The optimization-based method RobustFlow cannot work well, in which there are artifacts because fog degradation has violated the brightness constancy and gradient constancy assumption of optical flow. The semi-supervised method DenseFogFlow improves global motion smoothness, but suffers from severe outliers in motion boundaries in Fig. 7 (d). The reason is that DenseFogFlow only transfers motion knowledge from clean domain to synthetic foggy domain, but neglects the synthetic-to-real foggy domain gap, thus fails for real foggy scenes. Hardware-assisted GyroFlow can predict the accurate background optical flow, but the boundary of the foreground moving object is not clear in Fig. 7 (e). The reason is that the gyroscope data can only model the ego-motion, but it is difficult to estimate the independent foreground object motion under real foggy scenes. Instead, our method can remove erroneous outliers and obtain satisfactory results as shown in Fig. 7 (f).

3.3. Qualitative Comparisons on Real-Fog World

We also show the optical flow results of our proposed UCDA-Flow on our collected real foggy images, namely, Real-Fog World, in which the foreground and background motion patterns are more complex. The state-of-the-art optical flow methods in the comparisons on real foggy images include UFlow, RobustFlow, and DenseFogFlow. As shown in Fig. 8 (b), the unsupervised optical flow method UFlow cannot well predict optical flow under real foggy scenes because degradation breaks the brightness and gradient con-

stancy assumptions of optical flow. In Fig. 8 (c), we can observe that the RobustFlow could roughly predict the motion tendency. The semi-supervised method DenseFogFlow improves global motion smoothness, but the motion boundaries are not clear in Fig. 8 (d). On the contrary, our proposed UCDA-Flow can estimate satisfactory optical flow on Real-Fog World which has more complex motion patterns compared to the Fog-GOF dataset, as shown in Fig. 8 (e).



(a) Foggy images (b) w/o Cumulative DA (c) w/ DAMA, w/o CAMA (d) w/ Cumulative DA

Figure 1. Effectiveness of cumulative adaptation architecture. It is difficult to directly learn optical flow well from the real foggy domain without cumulative domain adaptation. DAMA can remove artifacts caused by fog degradation and make the optical flow clear, but there exist some mismatched regions. CAMA can further greatly refine the details of optical flow.

Figure 2. Effectiveness of depth association. Depth contains rich structure information, facilitating the optical flow rigid boundaries sharp.

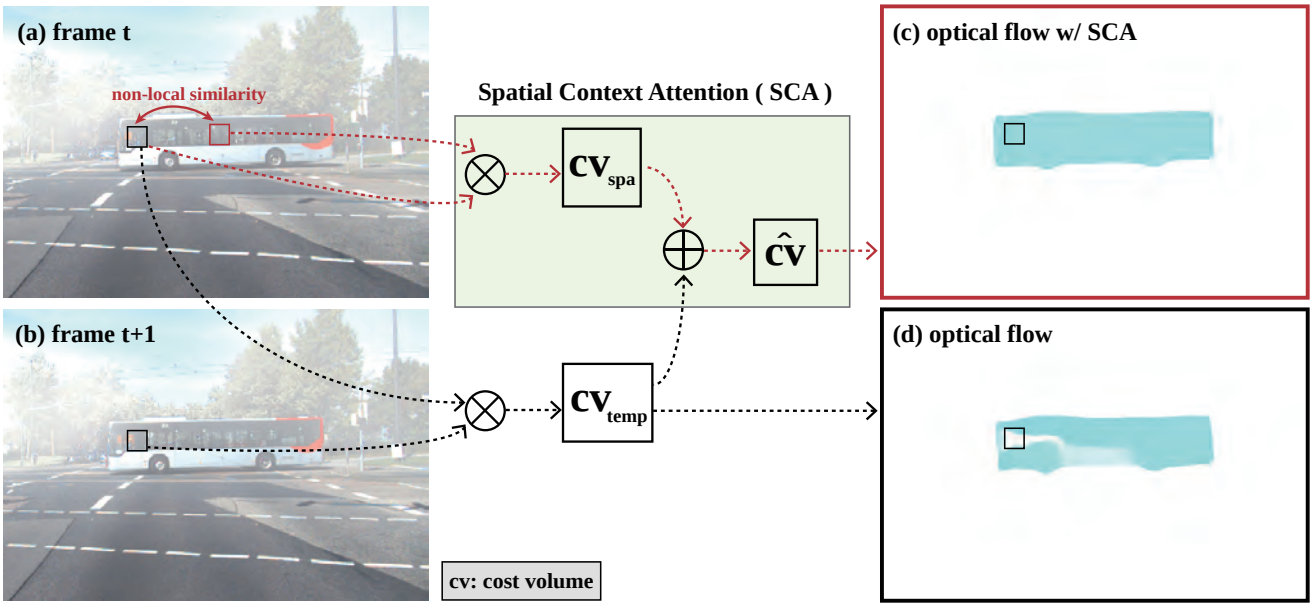


Figure 3. Effectiveness of spatial context attention module. In (d), there exist local outliers in optical flow obtained from temporal cost volume only in (d). In (c), **SCA** module can correct the optical flow details. **SCA** module can fuse the temporal cost volume with the spatial cost volume computed by the non-local similarity method, which is beneficial to refine optical flow.

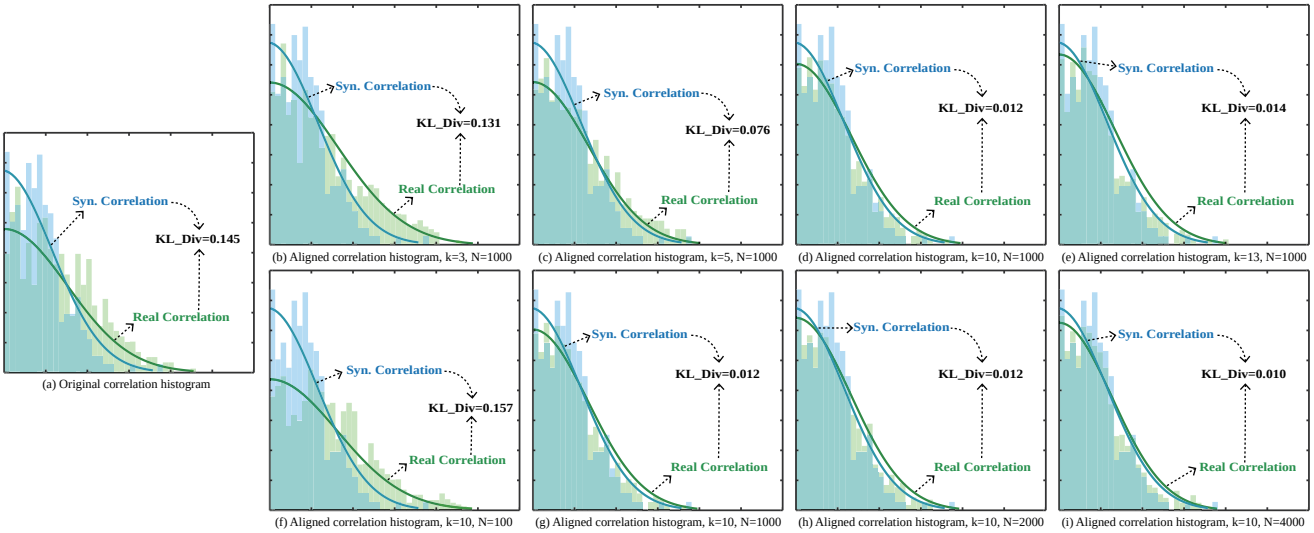


Figure 4. Ablation study on the main parameters of correlation distribution alignment module. Note that, the x-axis denotes the correlation value, and the y-axis denotes the number of features reaching the correlation. In (b) - (e), a large class parameter k can make correlation distributions close. In (f) - (i), the larger the sample number N, the lower the KL divergence, and the closer the correlation distributions.

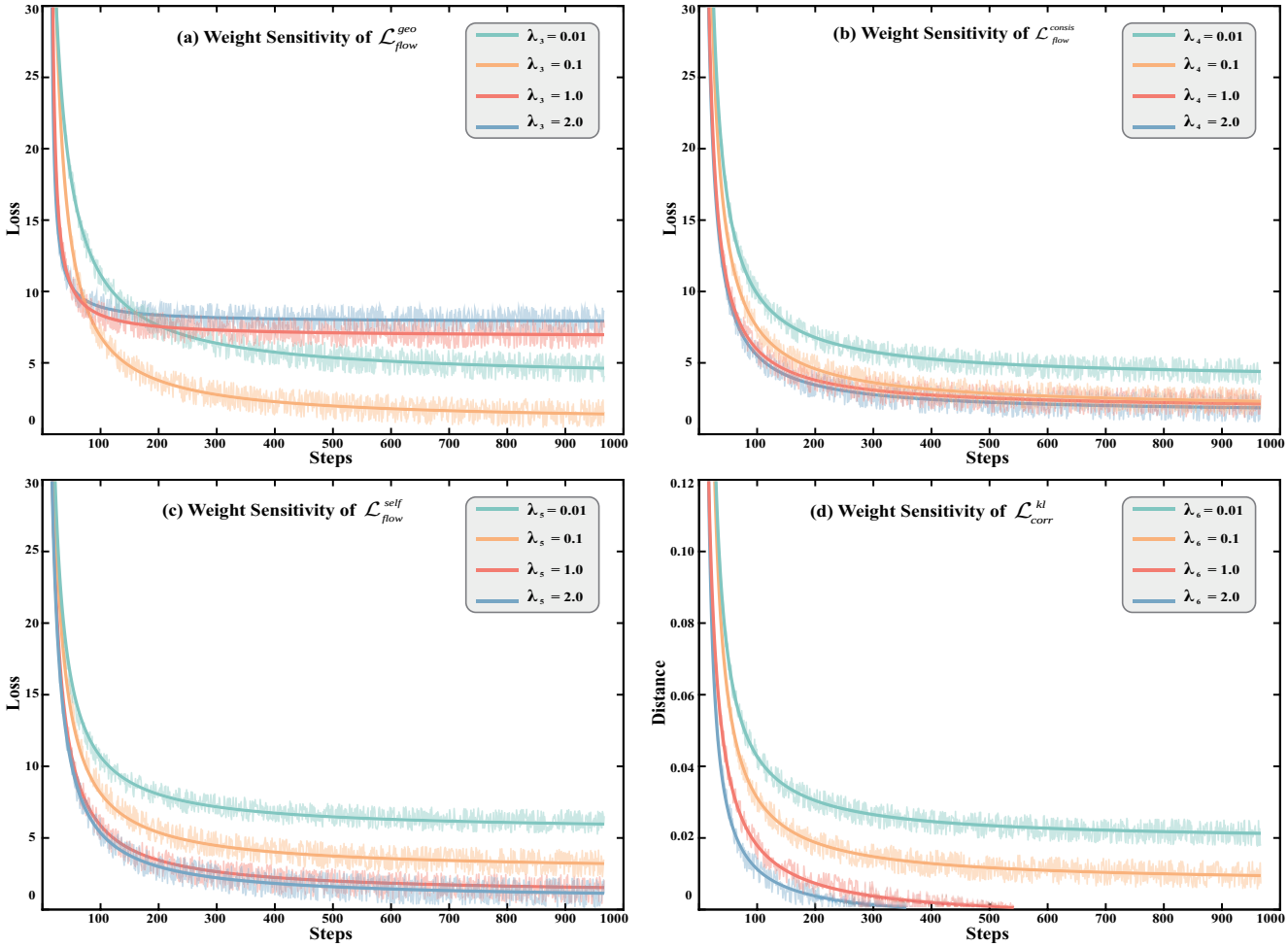


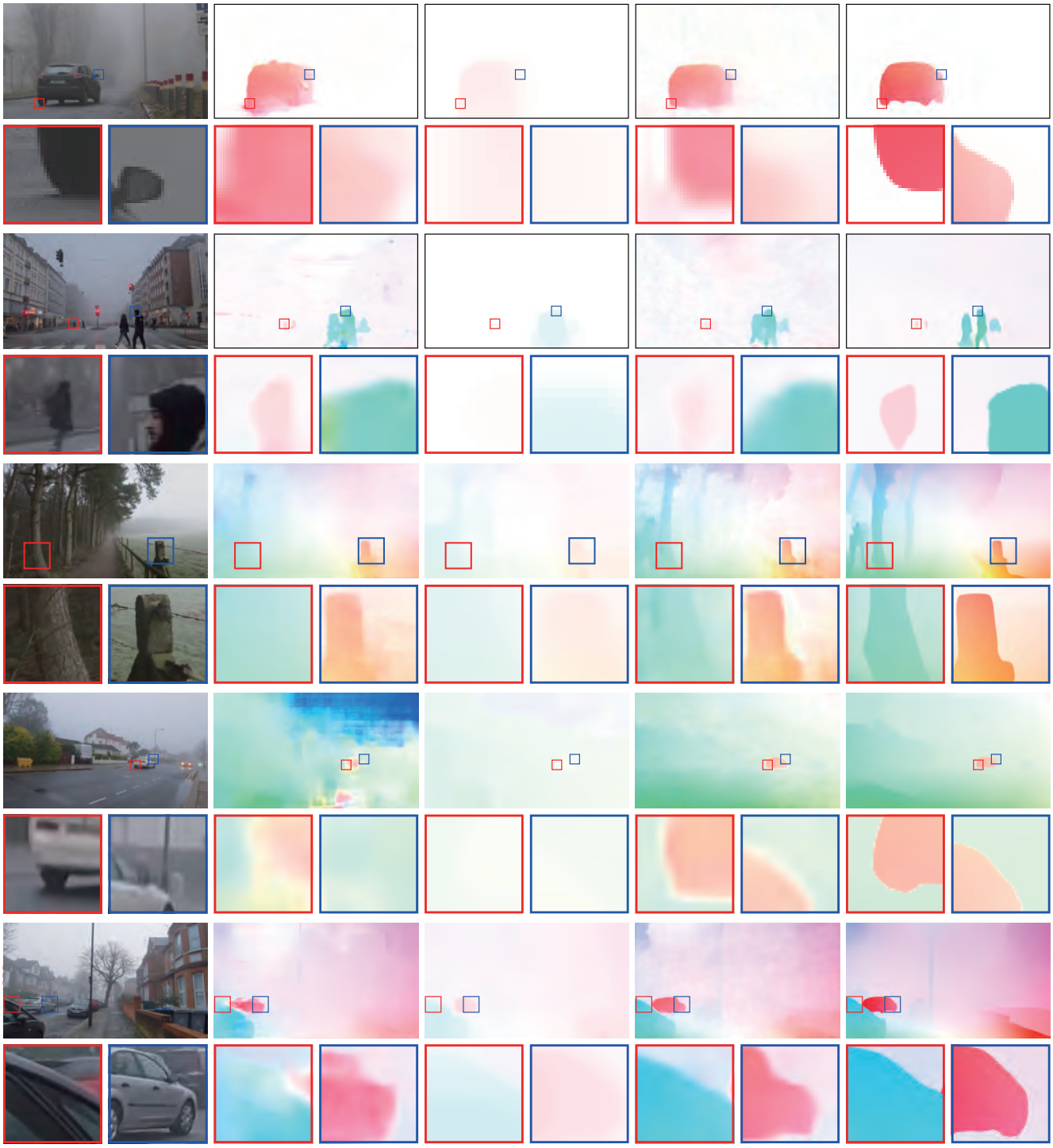
Figure 5. The weight sensitivity of model adaptation losses. The proposed framework is sensitive to the KL divergence loss of correlation distribution and the depth-association optical flow geometry loss.

(a) Foggy images / GT (b) RobustFlow (c) UFlow (d) Defog + UFlow (e) Defog + Selflow (f) UCDA-Flow

Figure 6. Comparison of optical flows on synthetic Fog-KITTI2015 dataset.

(a) Real foggy images (b) Groud Truth (c) RobustFlow (d) DenseFogFlow (e) GyroFlow (f) UCDA-Flow

Figure 7. Comparison of optical flows on real Fog-GOF dataset.



(a) Real foggy images (b) UFlow (c) RobustFlow (d) DenseFogFlow (e) UCDA-Flow

Figure 8. Comparison of optical flows on Real-Fog World dataset.