

Unsupervised Cumulative Domain Adaptation for Foggy Scene Optical Flow

Hanyu Zhou¹, Yi Chang^{1*}, Wending Yan², Luxin Yan¹

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

Abstract

Optical flow has achieved great success under clean scenes, but suffers from restricted performance under foggy scenes. To bridge the clean-to-foggy domain gap, the existing methods typically adopt the domain adaptation to transfer the motion knowledge from clean to synthetic foggy domain. However, these methods unexpectedly neglect the synthetic-to-real domain gap, and thus are erroneous when applied to real-world scenes. To handle the practical optical flow under real foggy scenes, in this work, we propose a novel unsupervised cumulative domain adaptation optical flow (UCDA-Flow) framework: depth-association motion adaptation and correlation-alignment motion adaptation. Specifically, we discover that depth is a key ingredient to influence the optical flow: the deeper depth, the inferior optical flow, which motivates us to design a depth-association motion adaptation module to bridge the clean-to-foggy domain gap. Moreover, we figure out that the cost volume correlation shares similar distribution of the synthetic and real foggy images, which enlightens us to devise a correlation-alignment motion adaptation module to distill motion knowledge of the synthetic foggy domain to the real foggy domain. Note that synthetic fog is designed as the intermediate domain. Under this unified framework, the proposed cumulative adaptation progressively transfers knowledge from clean scenes to real foggy scenes. Extensive experiments have been performed to verify the superiority of the proposed method.

1. Introduction

Optical flow has made great progress under clean scenes, but may suffer from restricted performance under foggy scenes [15]. The main reason is that fog weakens scene contrast, breaking the brightness and gradient constancy assumptions, which most optical flow methods rely on.

To alleviate this, researchers start from the perspective

*Corresponding author.

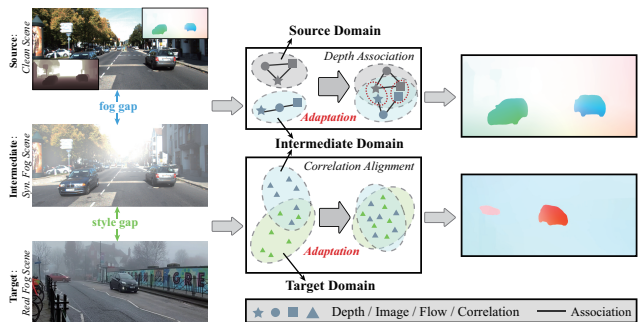


Figure 1. Illustration of the main idea. We propose to transfer motion knowledge from the source domain (clean scene) to the target domain (real foggy scene) through two-stage adaptation. We design the synthetic foggy scene as the intermediate domain. As for the clean-to-foggy domain gap (fog), we transfer motion knowledge from the source domain to the intermediate domain via depth association. As for the synthetic-to-real domain gap (style), we distill motion knowledge of the intermediate domain to the target domain by aligning the correlation of both the domains.

of domain adaptation, which mainly seeks the degradation-invariant features to transfer the motion knowledge from the clean scene to the adverse weather scene [14–16, 40]. For example, Li [15, 16] attempted to learn degradation-invariant features to enhance optical flow under rainy scenes in a supervised manner. Yan *et al.* [40] proposed a semi-supervised framework for optical flow under dense foggy scenes, which relies on the motion-invariant assumption between the paired clean and synthetic foggy images. These pioneer works have made a good attempt to handle the clean-to-foggy domain gap with synthetic degraded images through one-stage domain adaptation. However, they lack the constraints to guide the network to learn the motion pattern of real foggy domain, and fail for real foggy scenes. In other words, they have unexpectedly neglected the synthetic-to-real domain gap, thus limiting their performances on real-world foggy scenes. In this work, our goal is to progressively handle the two domain gaps: the clean-to-foggy gap and the synthetic-to-real gap in a cumulative domain adaptation framework in Fig. 1.

As for the clean-to-foggy gap, we discover that depth is a key ingredient to influence the optical flow: the deeper the depth, the inferior the optical flow. This observation inspires us to explore the usage of depth as the key to bridging the clean-to-foggy gap (seeing the fog gap in Fig. 1). On one hand, depth physically associates the clean image with the foggy image through atmospheric scattering model [26]; on the other hand, there exists a natural 2D-3D geometry projection relationship between depth and optical flow, which is used as a constraint to transfer motion knowledge from the clean domain to the synthetic foggy domain.

As for the synthetic-to-real gap, we figure out that cost volume correlation shares similar distribution of synthetic and real foggy images. Cost volume stores correlation value, which can physically measure the similarity between adjacent frames, regardless of synthetic and real foggy images. Therefore, cost volume benefits to bridging the synthetic-to-real domain gap (seeing the style gap in Fig. 1). We align the correlation distributions to distill motion knowledge of the synthetic foggy domain to the real foggy domain.

In this work, we propose a novel unsupervised cumulative domain adaptation optical flow (UCDA-Flow) framework for real foggy scene, including depth-association motion adaptation (DAMA) and correlation-alignment motion adaptation (CAMA). Specifically, in DAMA stage, we first estimate optical flow, ego-motion and depth with clean stereo images, and then project depth into optical flow space with 2D-3D geometry formula between ego-motion and scene-motion to enhance rigid motion. To bridge the clean-to-foggy gap, we utilize atmospheric scattering model [26] to synthesize the corresponding foggy images, and then transfer motion knowledge from the clean domain to the synthetic foggy domain. In CAMA stage, to bridge the synthetic-to-real domain gap, we transform the synthetic and real foggy images to the cost volume space, in which we align the correlation distribution to distill the motion knowledge of the synthetic foggy domain to the real foggy domain. The proposed cumulative domain adaptation framework could progressively transfer motion knowledge from clean domain to real foggy domain via depth association and correlation alignment. Overall, our main contributions are summarized as follows:

- We propose an unsupervised cumulative domain adaptation framework for optical flow under real foggy scene, consisting of depth-association motion adaptation and correlation-alignment motion adaptation. The proposed method can transfer motion knowledge from clean domain to real foggy domain through two-stage adaptation.
- We reveal that foggy scene optical flow deteriorates with depth. The geometry relationship between depth and optical flow motivates us to design a depth-association motion adaptation to bridge the clean-to-foggy domain gap.
- We illustrate that cost volume correlation distribution of the synthetic and real foggy images is consistent. This

prior benefits to close the synthetic-to-real domain gap through correlation-alignment motion adaptation.

2. Related Work

Optical Flow. Optical flow is the task of estimating per-pixel motion between video frames. Traditional methods [33] often formulate optical flow as an energy minimization problem. In recent years, the learning-based optical flow approaches [1, 2, 4, 6–8, 10, 13, 17, 23, 28, 30, 34, 35, 44, 45] have been proposed to improve the feature representation. PWC-Net [35] applied warp and cost volume to physically estimate optical flow in a coarse-to-fine pyramid. RAFT [36] was an important development of PWC-Net, which replaced the pyramid architecture with GRU [3] and constructed 4D cost volume for all pairs of pixels. To improve the motion feature representation, GMA [9] incorporated transformer into optical flow estimation and achieved better performance than RAFT. Furthermore, to relieve the dependency on synthetic datasets, the authors [11, 18, 24, 30, 42, 44, 47] proposed the unsupervised CNN optical flow methods with photometric loss or data distillation loss. Although they have achieved satisfactory results in clean scenes, they would suffer from degradation under foggy scenes. In this work, we propose an unsupervised optical flow framework for real foggy scenes.

Optical Flow under Adverse Weather. The robust optical flow estimation has been extensively studied for various adverse weather, such as rain [16], fog [40]. An intuitive solution to this challenging task is to perform the image deraining [5, 21, 41, 43, 46] or defogging [20, 27, 29, 31, 38] with subsequent optical flow estimation. However, existing derain/defog methods are not designed for optical flow and the possible over-smoothness or residual artifacts would contribute negative to optical flow. To bridge the clean-to-degraded gap, the authors [14–16, 40] have attempted to design the domain-invariant features to transfer motion knowledge from clean domain to synthetic degraded domain through one-stage adaptation. For example, Li *et al.* [15, 16] attempted to design rain-invariant features in a unified framework for robust optical flow under rainy scenes with synthetic degraded images. Yan *et al.* [40] estimated optical flow under dense foggy scenes via optical flow consistency. Li *et al.* [14] resorted to auxiliary gyroscope information which is robust to degradation for adverse weather optical flow. To further close the synthetic-to-real domain gap, we propose a two-stage cumulative domain adaptation framework for optical flow under real foggy scenes, which can bridge the clean-to-foggy and synthetic-to-real domain gaps.

3. Unsupervised Cumulative Adaptation

3.1. Overall Framework

The goal of this work is to estimate optical flow under real foggy scenes. Most existing adverse weather optical flow

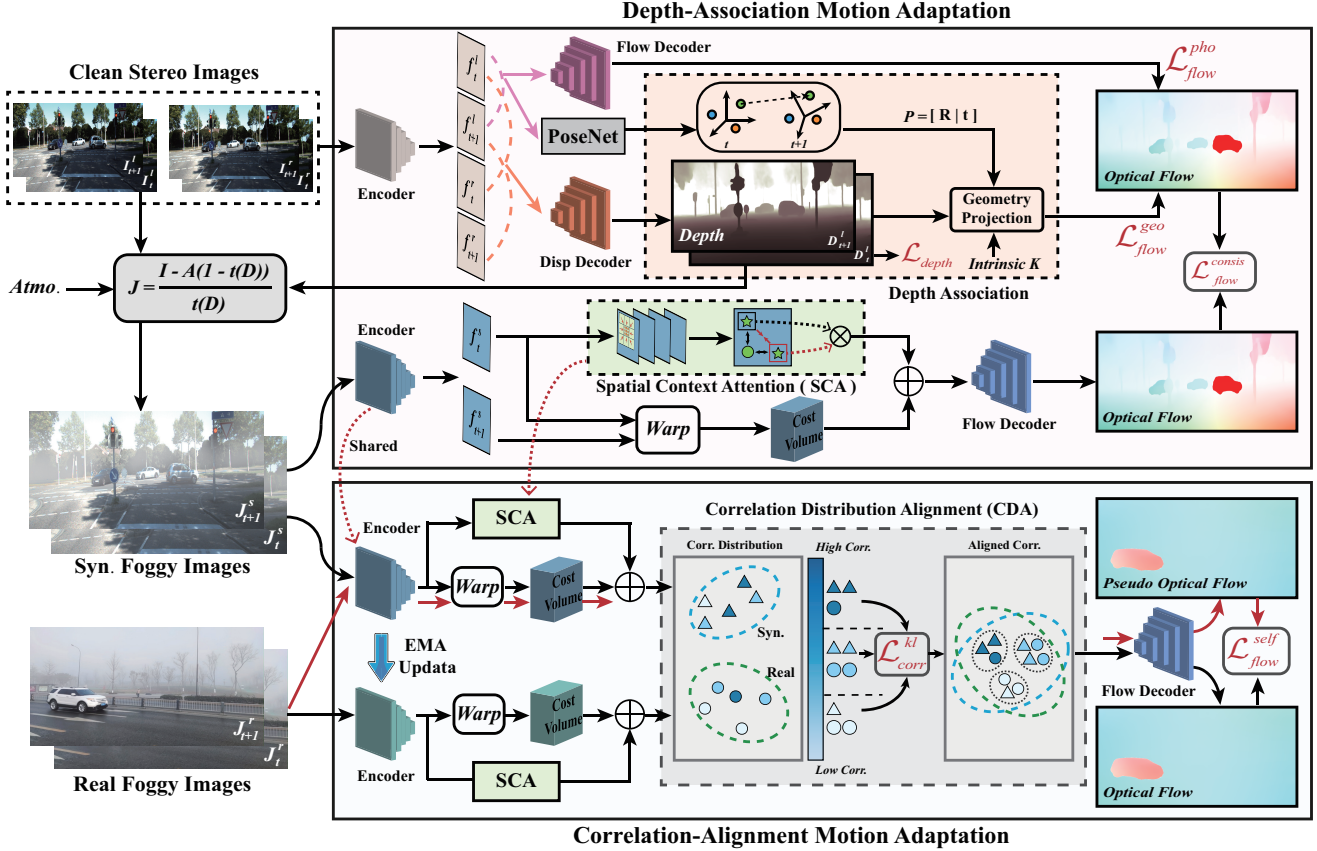


Figure 2. The architecture of the UCDA-Flow mainly contains depth-association motion adaptation (DAMA) and correlation-alignment motion adaptation (CAMA). The goal of DAMA stage is to bridge the clean-to-foggy domain gap, in which we associate the depth with optical flow via geometry projection and synthetic foggy images with atmospheric scattering model, and transfer motion knowledge from the clean domain to the synthetic foggy domain. In CAMA stage, to further close the synthetic-to-real domain gap, we align the correlation distributions of synthetic and real foggy images to distill motion knowledge of the synthetic foggy domain to the real foggy domain.

methods mainly adopt the one-stage adaptation to transfer motion knowledge from clean domain to synthetic adverse weather domain. However, due to the synthetic-to-real domain gap that these methods neglect, they cannot generalize well for real degraded scenes. In this work, we illustrate that foggy scene optical flow deteriorates with depth, which can bridge the clean-to-foggy domain gap. Moreover, we figure out that cost volume correlation shares the similar distribution of synthetic and real foggy images, benefiting to bridge the synthetic-to-real domain gap. Motivated by these analyses, we propose a novel unsupervised cumulative domain adaptation framework for optical flow under real foggy scenes. As shown in Fig. 2, our framework consists of two main modules: depth-association motion adaptation (DAMA) and correlation-alignment motion adaptation (CAMA). The DAMA associates depth with optical flow via geometry projection, renders synthetic foggy images with atmospheric scattering model, and transfers motion knowledge from clean domain to synthetic foggy domain. The CAMA aligns the correlation distribution of synthetic and real foggy images to distill motion knowledge of synthetic foggy do-

main to real foggy domain. Under this unified framework, the proposed framework could progressively transfer motion knowledge from clean domain to real foggy domain.

3.2. Depth-Association Motion Adaptation

The previous methods [16, 40] have attempted to directly transfer motion knowledge from clean domain to synthetic degraded domain. However, different from rain and snow, fog is a non-uniform degradation related to depth. This makes us naturally consider whether the optical flow affected by fog could be related to depth or not.

To illustrate this, we conduct an analysis experiment on the influence of fog on the image and optical flow along different depths in Fig. 3. We take clean KITTI2015 [25] and synthetic Fog-KITTI2015 as the experimental datasets. Compared to the corresponding clean images, we count the PSNR and the optical flow EPE of the foggy images at different depths. As the depth value becomes larger, the lower the PSNR, the higher the optical flow EPE, which means that the degradation of the image and optical flow aggravates with the larger depth. Moreover, we visualize the

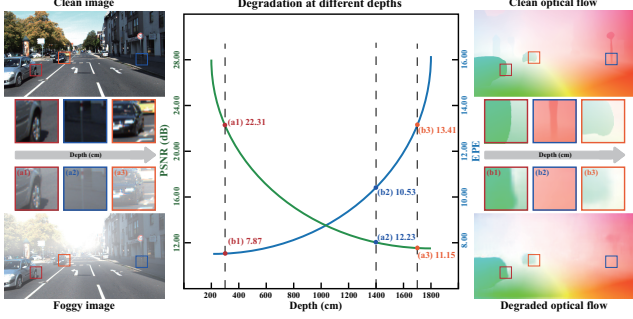


Figure 3. Analysis of fog degradation at different depths. The deeper the depth, the more severe the image and the optical flow. Depth is the key to bridging the clean-to-foggy domain gap.

images (Fig. 3 (a1)-(a3)) and optical flows (Fig. 3 (b1)-(b3)) at three depths. We can observe that the contrast of images and boundaries of optical flows become more blurry with the increasing of the depth. This inspires us that depth is the key to bridging the clean-to-foggy domain gap. On one hand, depth is associated with fog through atmospheric scattering model [26], which bridges the clean-to-foggy domain gap; on the other hand, depth could be used to refine optical flow of clean domain via strict geometry projection, and serve as a constraint to transfer motion knowledge from the clean domain to the synthetic foggy domain. Therefore, we propose a depth-association motion adaptation module to transfer motion knowledge between both the domains.

Depth Association. Given consecutive stereo images $[I_t^l, I_{t+1}^l, I_t^r, I_{t+1}^r]$, we first take RAFT [36] to estimate optical flow F with optical flow photometric loss [44] in an unsupervised manner as follow,

$$\mathcal{L}_{flow}^{pho} = \sum \psi(I_t^l - \text{warp}(I_{t+1}^l)) \odot (1 - O_f) / \sum (1 - O_f) + \sum \psi(I_{t+1}^l - \text{warp}(I_t^l)) \odot (1 - O_b) / \sum (1 - O_b), \quad (1)$$

where warp is the warping operator, ψ is a sparse L_p norm ($p = 0.4$). O_f and O_b are the forward and backward occlusion mask by checking forward-backward consistency, and \odot is a matrix element-wise multiplication. Similar to optical flow, stereo depths $[D_t^l, D_{t+1}^l]$ can be obtained by DispNet [39] via photometric loss and smooth loss,

$$\mathcal{L}_{depth} = \sum \psi(I_t^l - \text{warp}(I_{t+1}^l)) + |\nabla^2 D_t^l| e^{-|\nabla^2 I_t^l|} + \sum \psi(I_{t+1}^l - \text{warp}(I_t^l)) + |\nabla^2 D_{t+1}^l| e^{-|\nabla^2 I_{t+1}^l|}. \quad (2)$$

Here we wish to establish the dense pixel correspondence between the two adjacent frames through depth. Let p_t denotes the 2D homogeneous coordinate of a pixel in frame I_t^l and K denotes the camera intrinsic matrix. We can compute the corresponding point of p_t in frame I_{t+1}^l using the geometry projection equation [48],

$$p_{t+1} = KPD_t^l(p_t)K^{-1}p_t, \quad (3)$$

where P is the relative camera motion estimated by the pre-trained PoseNet [12]. We can then compute the rigid flow F_{rigid} at pixel p_t in I_t^l by, $F_{rigid}(p_t) = p_{t+1} - p_t$. We fur-

ther enhance motion in rigid regions with the consistency between the geometrically computed rigid flow and the directly estimated optical flow,

$$\mathcal{L}_{flow}^{geo} = \sum \|F - F_{rigid}\|_1 \odot (1 - V) / \sum (1 - V), \quad (4)$$

where V denotes the non-rigid region extracted from stereo clean images by forward-backward consistency check [49]. **Motion Knowledge Transfer.** To associate depth with fog, we synthesize the foggy images $[J_t^s, J_{t+1}^s]$ corresponding to the clean images using atmospheric scattering model [26],

$$J = \frac{I - A(1 - t(D))}{t(D)}, \quad (5)$$

where A denotes the predefined atmospheric light. $t(\cdot)$ is a decay function related depth. We then encode the synthetic foggy images into motion features $[f_t^s, f_{t+1}^s]$, and compute the temporal cost volume $cv_{temp} = (f_t^s)^T \cdot w(f_{t+1}^s)$, where T denotes transpose operator and w is the warp operator. Note that, in order to enable the flow model to have a suitable receptive field for smooth constraint of motion feature, we employ a spatial context attention (SCA) module with a non-local strategy [22]. Specifically, we devise a sliding window with a learnable kernel on the motion feature f_t^s to match the non-local similar feature f_{sim}^s , and generate k similar features corresponding to the cropped features from the motion feature f_t^s during sliding searching, as $[f_{sim}^1, f_{sim}^2, \dots, f_{sim}^k]$. And then we compute the spatial attention cost volume,

$$cv_{spa} = \frac{1}{k} \sum_{i=1}^k f_{sim}^i \cdot f_t^s. \quad (6)$$

The fused cost volume cv_s is produced by a residual operator as $\hat{cv}_s = cv_{temp} + \alpha cv_{spa}$, where α denotes a fusion weight. After that, we decode the fused cost volume to estimate optical flow F_{syn} of synthetic foggy images. We further transfer the pixel-wise motion knowledge from clean domain to synthetic foggy domain via flow consistency loss,

$$\mathcal{L}_{flow}^{consis} = \sum \|F_{syn} - F\|_1. \quad (7)$$

3.3. Correlation-Alignment Motion Adaptation

Although depth-association motion adaptation can bridge the clean-to-foggy domain gap and provide a coarse optical flow for synthetic foggy domain, it cannot help the synthetic-to-real domain gap. Hence, our method may inevitably suffer from artifacts under real foggy scenes due to the synthetic-to-real domain gap. To explore how large the synthetic-to-real foggy domain gap is, we visualize the feature distributions of the synthetic and real foggy images via t-SNE [37] in Fig. 4 (a). The degradation pattern discrepancy between synthetic and real foggy images is small, but there exists an obvious synthetic-to-real style gap that restricts the optical flow performance under real foggy scenes.

Direct motion adaptation from synthetic to real domain is difficult, since their background is different. Our solution is to construct an intermediate domain as an adaptation bridge namely cost volume, physically measuring the similarity between adjacent frames, not limited by scene difference. We

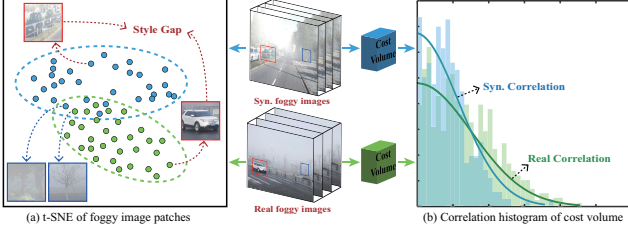


Figure 4. Visual distribution of synthetic and real foggy images. In (a) t-SNE of foggy image patches, the difference of degradation pattern is small, but there exists an obvious style gap between both the domains. In (b) correlation histogram of cost volume, the entire correlation distributions are similar. This motivates us to close the synthetic-to-real domain gap with correlation distribution.

transform foggy images to cost volume space, and visualize the correlation distributions of synthetic and real foggy images via the histogram in Fig. 4 (b). We can observe that both the domains share a similar correlation distribution. This motivates us to provide a novel correlation-alignment motion adaptation module, which can distill motion knowledge of the synthetic foggy domain to the real foggy domain.

Correlation Distribution Alignment (CDA). We begin with two encoder E_s, E_r for the synthetic foggy images $[J_t^s, J_{t+1}^s]$ and the real foggy images $[J_t^r, J_{t+1}^r]$, respectively. We encode them to obtain the cost volume cv_s, cv_r with the warp operator and the SCA module. Furthermore, we randomly sample N correlation values in the cost volumes cv_s, cv_r to represent the entire correlation distribution of cost volume normalized into $[0, 1]$ for both the domains. According to the range of correlation, we choose threshold values $[\delta_1, \delta_2, \dots, \delta_{k-1}]$ to label the sampled correlation into k classes, such as high correlation, and low correlation. Then, the correlation distribution p is estimated by,

$$p = \frac{n+1}{N+k}, \quad (8)$$

where n is the number of the sampled correlation of one category. Note that we add an offset 1 to each category sampled correlation of Eq. 8 to ensure at least a single instance could be present in the real foggy domain. Thus, to align the features of the synthetic and real foggy domains, we minimize the correlation distribution distance between the two domains by enforcing Kullback-Leibler divergence,

$$\mathcal{L}_{corr}^{kl} = \sum_{i=1}^k p_{r,i} \log \frac{p_{r,i}}{p_{s,i}}, \quad (9)$$

where $p_{s,i}, p_{r,i}$ denote the i category sampled correlation distributions of the synthetic foggy domain and the real foggy domain, respectively. The aligned correlation distributions represent that both the domains could have similar optical flow estimation capabilities. Finally, we decode the aligned cost volume to predict optical flow for real foggy images.

Self-Supervised Training Strategy. To improve the robustness of knowledge transfer, we present a self-supervised training strategy that attempts to transfer motion knowledge

from the synthetic foggy domain to the real foggy domain at the optical flow field level. We feed the real foggy images $[J_t^r, J_{t+1}^r]$ to the flow network of the synthetic foggy domain, which outputs the optical flow as the pseudo-labels F_{pseudo} (seeing the red arrow in Fig. 2). We then impose a self-supervised loss on the optical flow F_{real} estimated by the flow network of the real foggy domain,

$$\mathcal{L}_{flow}^{self} = \sum \|F_{real} - F_{pseudo}\|_1. \quad (10)$$

During the training process, the encoder $E_r(f; \theta_r)$ of the real foggy domain is updated with the encoder $E_s(f; \theta_s)$ of the synthetic foggy domain using the exponential moving average (EMA) mechanism, namely, $\theta_r \leftarrow \theta_r \cdot \lambda + \theta_s \cdot (1 - \lambda)$, where λ controls the window of EMA and is often close to 1.0. The proposed correlation-alignment motion adaptation distills motion knowledge of the synthetic foggy domain to the real foggy domain in the feature correlation and optical flow dimensions, respectively.

3.4. Total Loss and Implementation Details

Consequently, the total objective for the proposed framework is written as follows,

$$\begin{aligned} \mathcal{L} = & \lambda_1 \mathcal{L}_{depth} + \lambda_2 \mathcal{L}_{flow}^{pho} + \lambda_3 \mathcal{L}_{flow}^{geo} \\ & + \lambda_4 \mathcal{L}_{flow}^{consis} + \lambda_5 \mathcal{L}_{flow}^{self} + \lambda_6 \mathcal{L}_{corr}, \end{aligned} \quad (11)$$

where the first four terms are the unsupervised losses that aim to transfer knowledge from the clean domain to the synthetic foggy domain, and the intention of the last two terms is to build the mathematical relationship between synthetic and real foggy domains. We empirically set the parameters $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6\} = \{1, 1, 0.1, 1, 1, 0.1\}$. Besides, as for the parameters of CDA, we set the sample number N as 1000 and the number of categories k as 10. The classification threshold values δ are set linearly from $[0, 1]$. The weight λ of the EMA for self-supervised training strategy is 0.99.

The proposed framework UCDA-Flow consists of three encoders, two flow decoders, one disp decoder, and one residual block for SCA. We first train the optical flow network and the disp network of the clean domain via \mathcal{L}_{flow}^{pho} , \mathcal{L}_{depth} and \mathcal{L}_{flow}^{geo} . We update the optical flow network of the synthetic foggy domain via $\mathcal{L}_{flow}^{consis}$. Then we update the optical flow network of the real foggy domain via $\mathcal{L}_{flow}^{self}$ and \mathcal{L}_{corr}^{kl} with 1000 epochs and 0.0005 learning rate. After that, we optimize the whole framework via the full loss \mathcal{L} with 0.0002 learning rate. At the test stage, the testing model only needs the optical flow network of the real foggy domain, including encoder, warp, cost volume, and flow decoder.

4. Experiments

4.1. Experiment Setup

Dataset. We take the KITTI2015 [25] dataset as the representative clean scene. We validate the performance of optical

Table 1. Quantitative results on synthetic Light Fog-KITTI2015 (LF-KITTI) and Dense Fog-KITTI2015 (DF-KITTI) datasets.

Method	RobustFlow	DenseFogFlow	UFlow			Selflow			SMURF	UCDA-Flow	
			-	FFA-Net +	AECR-Net +	-	FFA-Net +	AECR-Net +			
LF-KITTI	EPE	23.48	6.82	14.33	14.21	11.66	13.42	13.15	10.06	10.48	5.94
	F1-all	81.54%	39.18%	56.96%	56.38%	50.92%	55.37%	54.83%	48.74%	47.60%	34.11%
DF-KITTI	EPE	25.32	8.03	16.55	15.97	12.16	15.84	14.93	11.21	11.56	6.29
	F1-all	85.77%	41.73%	62.84%	61.69%	53.17%	58.81%	57.06%	50.25%	51.39%	36.25%

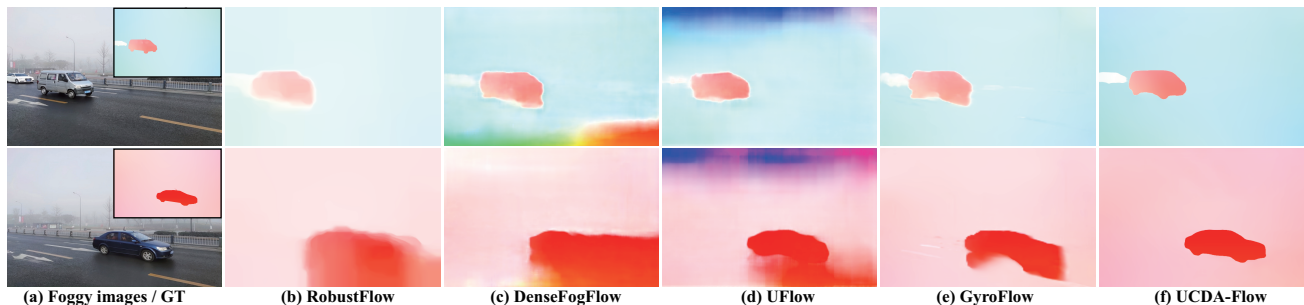


Figure 5. Visual comparison of optical flows on real Fog-GOF dataset.

Table 2. Quantitative results on real foggy datasets.

Method	Robust Flow	UFlow	GMA		Dense FogFlow	Gyro Flow	Ours	
			-	ssl				
Fog-GOF	EPE	12.25	2.97	1.63	1.69	1.78	0.95	0.81
	F1-all	80.93%	30.82%	14.25%	15.11%	16.41%	9.13%	7.18%
Dense-Fog	EPE	13.48	6.21	3.68	3.81	4.32	-	2.94
	F1-all	79.31%	62.45%	33.18%	35.20%	41.26%	-%	28.67%

flow on one synthetic and three real foggy datasets.

- **Fog-KITTI2015.** We construct a synthetic foggy KITTI dataset with different densities of fog (e.g., dense fog and light fog) onto images of KITTI2015 [25] via atmospheric scattering model [26]. We select 8400 images of Fog-KITTI2015 dataset for training and 400 images for testing.
- **Fog-GOF.** GOF [14] is a dataset containing four different scenes with synchronized gyro readings, such as regular scenes and adverse weather scenes. We choose foggy images of GOF to compose a new foggy dataset, namely Fog-GOF, of which 1000 images for training and 105 images for testing.
- **DenseFog.** We seek the real foggy dataset collected by DenseFogFlow [40], namely DenseFog, of which 2346 images for training and 100 images for testing.
- **Real-Fog World.** We collect degraded videos under real foggy scenes from *Youtube*, with 1200 and 240 images for training and testing, respectively.

Comparison Methods. We choose three competing methods GyroFlow [14], DenseFogFlow [40] and RobustFlow [15] which are designed for adverse weather optical flow. Moreover, we select several state-of-the-art supervised (GMA [9]) and unsupervised (SMURF [32], UFlow [11] and Selflow [19]) optical flow approaches designed for clean scenes. The unsupervised methods are first trained on KITTI2015 for initialization and re-trained on the target degraded dataset. The supervised method is first trained on

the synthetic dataset, and then trained on target real datasets via self-supervised learning [32], denoted with ‘ssl’ in Table 5. As for the comparison on Fog-KITTI2015, we design two different training strategies for competing methods. The first is that we directly train the comparison methods on foggy images. The second is to perform the defogging first via defog approaches (e.g., FFA-Net [27] and AECR-Net [38]), and then we train the comparison methods on the defogging results (named as FFA-Net+ / AECR-Net+).

Evaluation Metrics. We choose average endpoint error (EPE [4]) and the lowest percentage of flow outliers (F1-all [25]) as evaluation metrics for the quantitative evaluation. The smaller the index is, the better the predicted result is.

4.2. Experiments on Synthetic Images

In Table 1, we show the quantitative comparison of the synthetic light and dense Fog-KITTI2015 datasets. Note that, we choose unsupervised methods for fair comparison which all do not need any ground truth. We have two key observations. First, the proposed UCDA-Flow is significantly better than the unsupervised counterparts under light and dense foggy conditions. Since degradation breaks the basic assumption of optical flow, the competing methods cannot work well. Second, the pre-processing procedure defogging (e.g., FFA-Net / AECR-Net + UFlow) is positive to optical flow estimation. However, since the defog methods are not designed for optical flow and the defogging images may be over-smoothness, the performance of optical flow is still limited. On the contrary, the proposed method could well handle both light and dense foggy images. The reason is that the proposed UCDA-Flow bypasses the difficulties of directly estimating optical flow from degraded images, and transferring motion knowledge from clean domain to foggy domain via unsupervised domain adaptation.



Figure 6. Visual comparison of optical flows on Real-Fog World.

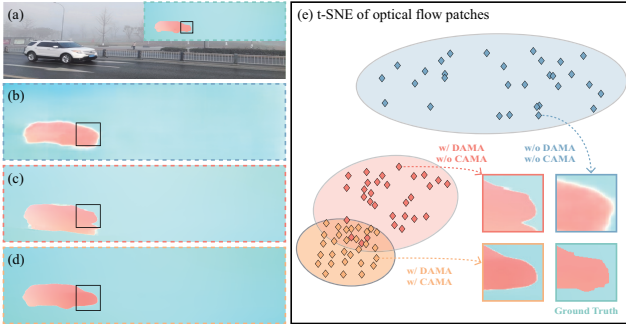


Figure 7. Effectiveness of cumulative adaptation architecture. (a) Real foggy image. (b)-(d) Optical flows without DA, with DAMA only, with DAMA and CAMA, respectively. (e) The t-SNE visualization of each adaptation strategy.

4.3. Experiments on Real Images

In Table 2, the quantitative results on Fog-GOF and DenseFog verify the superiority of our method. Note that, the performance barely changes before and after ‘ssl’. The reason is that fog has unexpectedly broken the photometric constancy assumption which self-supervised learning optical flow relies on, thus limiting its capability. In Fig. 5 and 6, we also show the visual comparison results on Fog-GOF and Real-Fog World datasets. The optimization-based method RobustFlow and the unsupervised method UFlow cannot work well. The supervised method GMA could obtain relatively smooth visualization results, but there exist outliers in Fig. 6 (e). The hardware-assisted GyroFlow heavily relies on the camera ego-motion captured by the gyroscope data yet is less effective for the independent foreground object motion in Fig. 5 (e). DenseFogFlow only bridges the clean-to-foggy domain gap, but neglects the synthetic-to-real foggy domain gap, thus suffers artifacts when applied to real foggy images in Fig. 5 and 6 (c). On the contrary, the proposed cumulative adaptation framework can obtain satisfactory results under real foggy scenes in Fig. 5 and 6 (f).

4.4. Ablation Study

Effectiveness of Cumulative Adaptation Architecture. To illustrate the effectiveness of cumulative DAMA-CAMA ar-

Table 3. Ablation study on adaptation losses.

$\mathcal{L}_{flow}^{consis}$	\mathcal{L}_{flow}^{geo}	$\mathcal{L}_{flow}^{self}$	\mathcal{L}_{corr}^{kl}	Fog-GOF	
				EPE	F1-all
×	×	×	×	2.92	30.94%
×	✓	×	×	2.88	30.20%
✓	×	×	×	1.59	14.03%
✓	✓	×	×	1.35	11.27%
✓	✓	✓	×	1.27	10.76%
✓	✓	×	✓	0.92	8.81%
✓	✓	✓	✓	0.81	7.18%

chitecture, in Fig. 7, we show the optical flow estimation of different adaptation strategies and visualize their low-dimensional distributions via t-SNE. In Fig. 7 (b), we can observe that there exist artifacts in the motion boundary without domain adaptation. With DAMA only in Fig. 7 (c), most of the outliers caused by degradation are removed. with both DAMA and CAMA in Fig. 7 (d), the motion boundary is clearer. Moreover, we visualize their corresponding t-SNE distribution in Fig. 7 (e). The blue, red, and yellow diamonds denote the distributions without domain adaptation, with DAMA only and with DAMA-CAMA, respectively. The blue distribution is scattered, the red distribution is gradually focused, and the yellow distribution is most concentrated, illustrating that the cumulative domain adaptation could progressively improve real foggy scene optical flow.

Effectiveness of Adaptation Losses. We study how the adaptation losses of the proposed method contribute to the final result as shown in Table 3. \mathcal{L}_{flow}^{geo} aim to enforce the optical flow in rigid regions. $\mathcal{L}_{flow}^{consis}$ is to transfer motion knowledge from the clean domain to the synthetic foggy domain. The goal of $\mathcal{L}_{flow}^{self}$ and \mathcal{L}_{corr}^{kl} is to distill motion knowledge of the synthetic foggy domain to the real foggy domain. We can observe that the motion consistency loss $\mathcal{L}_{flow}^{consis}$ make a major contribution to the optical flow result, and the correlation distribution alignment loss \mathcal{L}_{corr}^{kl} can further improve the optical flow under real foggy scenes.

4.5. Discussion

How dose the Depth Improve Optical Flow? We study the importance of depth in transferring motion knowledge

Table 4. The effect of modules in CAMA stage on optical flow.

EMA	SCA	CDA	Fog-GOF	
			EPE	F1-all
×	×	×	1.38	12.06%
✓	×	×	1.36	11.43%
✓	✓	×	1.27	10.76%
✓	✓	✓	0.81	7.18%

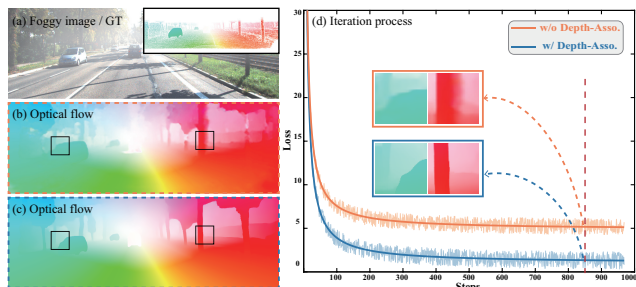


Figure 8. Effect of depth association on optical flow. (a) Foggy image. (b)-(c) Optical flows without depth-association and with depth-association, respectively. (d) Iteration process. Depth geometry association can improve the rigid motion boundary.

from clean domain to synthetic foggy domain in Fig. 8. Without depth association in Fig. 8 (b), the rigid motion boundaries are blurry. With depth association in Fig. 8 (c), the optical flow is global-smooth with sharp boundaries. Besides, we visualize their training iteration process in Fig. 8 (d). We can observe that depth association can further improve the optimal value that the proposed model converges to. Therefore, the depth geometry association can enhance the rigid motion boundary for synthetic foggy images.

Importance of Correlation Distribution Alignment. In Table 4, we show the effect of different modules on the optical flow of real foggy domain. **EMA** is to prevent the weights of the network from falling into the local optimum at the training stage. **SCA** aims to enhance the motion saliency in the cost volume. **CDA** is to transfer motion knowledge from the synthetic foggy domain to the real foggy domain by aligning the correlation distributions of both the domains. As shown in Table 4, **EMA** and **SCA** contribute a small improvement on the optical flow, while the **CDA** plays a key role in improving the optical flow of real foggy domain.

Why Associate Depth with Fog? We also study the effect of different foggy image synthesis strategies on the optical flow in Table 5. The GAN-based strategy cannot perform well. The reason is that GAN may erratically produce some new artifacts during the image translation, but instead exacerbate the synthetic-to-real foggy domain gap. On the contrary, since fog is a non-uniform degradation related to depth, it is reasonable that we use depth to synthesize foggy images. Note that, the depth estimated by monocular is not accurate enough due to the weak constraints. We also upsample the sparse depth in KITTI dataset into the dense depth to

Table 5. Choice of different foggy image synthesis strategies.

Method	Fog-GOF	
	EPE	F1-all
GAN-Based	1.43	13.10%
Depth-Based	Monocular	0.92 8.83%
	Pseudo-GT	0.83 7.45%
	Stereo (Ours)	0.81 7.18%

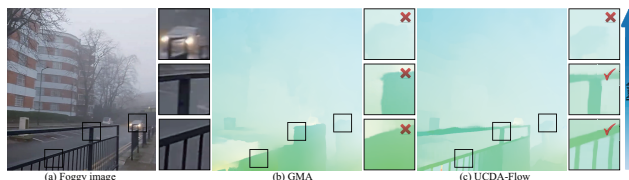


Figure 9. Limitation of the proposed method. Compared with the state-of-the-art optical flow method GMA [9], UCDA-Flow obtains the clear motion boundary in the nearby regions, but fails for the too-distant moving objects under foggy scenes.

synthesize the foggy images (pseudo-GT strategy), while this strategy is slightly inferior to our stereo-based strategy. The proposed depth association motion adaptation could make a positive contribution to motion knowledge transfer.

Limitation. In Fig. 9, we discuss the limitation of the proposed UCDA-Flow. Compared with the SOTA optical flow method GMA [9], UCDA-Flow obtains the clearer motion boundary in the too-distantons, but fails for the too distant moving objects under foggy scenes. There are two reasons for this problem. First, our framework requires depth to enhance the optical flow in rigid regions, but it is difficult for the stereo strategy to obtain accurate depth in distant regions. Second, degradation is so severe that the details of the distant moving object are lost. In the future, we attempt to employ lidar for detecting distant objects.

5. Conclusion

In this work, we propose an unsupervised cumulative domain adaptation framework for optical flow under real foggy scenes. We reveal that depth is a key ingredient to influence optical flow, which motivates us to design a depth-association motion adaptation to close the clean-to-foggy domain gap. We figure out that cost volume correlation shares a similar distribution of the synthetic and real foggy images, which enlightens us to devise a correlation-alignment motion adaptation to bridge the synthetic-to-real domain gap. We have conducted experiments on the synthetic and real foggy datasets to verify the superiority of our method.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grant 61971460, in part by JCJQ Program under Grant 2021-JCJQ-JJ-0060, in part by the National Natural Science Foundation of China under Grant 62101294, and in part by Xiaomi Young Talents Program.

References

- [1] Filippo Aleotti, Matteo Poggi, and Stefano Mattoccia. Learning optical flow from still images. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 15201–15211, 2021. [2](#)
- [2] Cheng Chi, Qingjie Wang, Tianyu Hao, Peng Guo, and Xin Yang. Feature-level collaboration: Joint unsupervised learning of optical flow, stereo depth and camera motion. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2463–2473, 2021. [2](#)
- [3] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014. [2](#)
- [4] 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 *Int. Conf. Comput. Vis.*, pages 2758–2766, 2015. [2](#), [6](#)
- [5] Xueyang Fu, Jiabin Huang, Xinghao Ding, Yinghao Liao, and John Paisley. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Trans. Image Process.*, 26(6):2944–2956, 2017. [2](#)
- [6] Zhaoyang Huang, Xiaoyu Shi, Chao Zhang, Qiang Wang, Ka Chun Cheung, Hongwei Qin, Jifeng Dai, and Hongsheng Li. Flowformer: A transformer architecture for optical flow. In *Eur. Conf. Comput. Vis.*, pages 668–685. Springer, 2022. [2](#)
- [7] Tak-Wai Hui, Xiaoou Tang, and Chen Change Loy. LiteFlowNet: A lightweight convolutional neural network for optical flow estimation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 8981–8989, 2018. [2](#)
- [8] Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox. FlowNet 2.0: Evolution of optical flow estimation with deep networks. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2462–2470, 2017. [2](#)
- [9] Shihao Jiang, Dylan Campbell, Yao Lu, Hongdong Li, and Richard Hartley. Learning to estimate hidden motions with global motion aggregation. In *Int. Conf. Comput. Vis.*, pages 9772–9781, 2021. [2](#), [6](#), [8](#)
- [10] Shihao Jiang, Yao Lu, Hongdong Li, and Richard Hartley. Learning optical flow from a few matches. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 16592–16600, 2021. [2](#)
- [11] Rico Jonschkowski, Austin Stone, Jonathan T Barron, Ariel Gordon, Kurt Konolige, and Anelia Angelova. What matters in unsupervised optical flow. In *Eur. Conf. Comput. Vis.*, pages 557–572. Springer, 2020. [2](#), [6](#)
- [12] Alex Kendall, Matthew Grimes, and Roberto Cipolla. PoseNet: A convolutional network for real-time 6-dof camera relocalization. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2938–2946, 2015. [4](#)
- [13] Wei-Sheng Lai, Jia-Bin Huang, and Ming-Hsuan Yang. Semi-supervised learning for optical flow with generative adversarial networks. *Adv. Neural Inform. Process. Syst.*, 30, 2017. [2](#)
- [14] Haipeng Li, Kunming Luo, and Shuaicheng Liu. Gyroflow: Gyroscope-guided unsupervised optical flow learning. pages 12869–12878, 2021. [1](#), [2](#), [6](#)
- [15] Ruoteng Li, Robby T Tan, and Loong-Fah Cheong. Robust optical flow in rainy scenes. In *Eur. Conf. Comput. Vis.*, pages 288–304, 2018. [1](#), [2](#), [6](#)
- [16] Ruoteng Li, Robby T Tan, Loong-Fah Cheong, Angelica I Aviles-Rivero, Qingnan Fan, and Carola-Bibiane Schonlieb. Rainflow: Optical flow under rain streaks and rain veiling effect. In *Int. Conf. Comput. Vis.*, pages 7304–7313, 2019. [1](#), [2](#), [3](#)
- [17] Liang Liu, Jiangning Zhang, Ruifei He, Yong Liu, Yabiao Wang, Ying Tai, Donghao Luo, Chengjie Wang, Jilin Li, and Feiyue Huang. Learning by analogy: Reliable supervision from transformations for unsupervised optical flow estimation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 6489–6498, 2020. [2](#)
- [18] Pengpeng Liu, Irwin King, Michael R Lyu, and Jia Xu. DdfFlow: Learning optical flow with unlabeled data distillation. In *AAAI Conf. on Arti. Intell.*, pages 8770–8777, 2019. [2](#)
- [19] Pengpeng Liu, Michael Lyu, Irwin King, and Jia Xu. SelfFlow: Self-supervised learning of optical flow. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 4571–4580, 2019. [6](#)
- [20] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. GridDehazenet: Attention-based multi-scale network for image dehazing. In *Int. Conf. Comput. Vis.*, pages 7314–7323, 2019. [2](#)
- [21] Yang Liu, Ziyu Yue, Jinshan Pan, and Zhixun Su. Unpaired learning for deep image deraining with rain direction regularizer. In *Int. Conf. Comput. Vis.*, pages 4753–4761, 2021. [2](#)
- [22] Ao Luo, Fan Yang, Xin Li, and Shuaicheng Liu. Learning optical flow with kernel patch attention. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 8906–8915, 2022. [4](#)
- [23] Kunming Luo, Chuan Wang, Shuaicheng Liu, Haoqiang Fan, Jue Wang, and Jian Sun. UpFlow: Upsampling pyramid for unsupervised optical flow learning. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1045–1054, 2021. [2](#)
- [24] Simon Meister, Junhwa Hur, and Stefan Roth. UnFlow: Unsupervised learning of optical flow with a bidirectional census loss. In *AAAI Conf. on Arti. Intell.*, pages 7251–7259, 2018. [2](#)
- [25] Moritz Menze and Andreas Geiger. Object scene flow for autonomous vehicles. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3061–3070, 2015. [3](#), [5](#), [6](#)
- [26] Srinivasa G Narasimhan and Shree K Nayar. Vision and the atmosphere. *Int. J. Comput. Vis.*, 48(3):233–254, 2002. [2](#), [4](#), [6](#)
- [27] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *AAAI Conf. on Arti. Intell.*, pages 11908–11915, 2020. [2](#), [6](#)
- [28] Anurag Ranjan and Michael J Black. Optical flow estimation using a spatial pyramid network. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 4161–4170, 2017. [2](#)
- [29] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3253–3261, 2018. [2](#)

- [30] Zhe Ren, Junchi Yan, Bingbing Ni, Bin Liu, Xiaokang Yang, and Hongyuan Zha. Unsupervised deep learning for optical flow estimation. In *AAAI Conf. on Arti. Intell.*, pages 1495–1501, 2017. 2
- [31] Yuanjie Shao, Lerenhan Li, Wenqi Ren, Changxin Gao, and Nong Sang. Domain adaptation for image dehazing. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2808–2817, 2020. 2
- [32] Austin Stone, Daniel Maurer, Alper Ayvaci, Anelia Angelova, and Rico Jonschkowski. Smurf: Self-teaching multi-frame unsupervised raft with full-image warping. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3887–3896, 2021. 6
- [33] Deqing Sun, Stefan Roth, and Michael J Black. Secrets of optical flow estimation and their principles. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2432–2439, 2010. 2
- [34] Deqing Sun, Daniel Vlasic, Charles Herrmann, Varun Jampani, Michael Krainin, Huiwen Chang, Ramin Zabih, William T Freeman, and Ce Liu. Autoflow: Learning a better training set for optical flow. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 10093–10102, 2021. 2
- [35] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 8934–8943, 2018. 2
- [36] Zachary Teed and Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow. In *Eur. Conf. Comput. Vis.*, pages 402–419, 2020. 2, 4
- [37] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. 9(11), 2008. 4
- [38] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. pages 10551–10560, 2021. 2, 6
- [39] Haofei Xu and Juyong Zhang. Aanet: Adaptive aggregation network for efficient stereo matching. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1959–1968, 2020. 4
- [40] Wending Yan, Aashish Sharma, and Robby T Tan. Optical flow in dense foggy scenes using semi-supervised learning. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 13259–13268, 2020. 1, 2, 3, 6
- [41] Wending Yan, Robby T Tan, Wenhan Yang, and Dengxin Dai. Self-aligned video deraining with transmission-depth consistency. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 11966–11976, 2021. 2
- [42] Gengshan Yang and Deva Ramanan. Volumetric correspondence networks for optical flow. *Adv. Neural Inform. Process. Syst.*, 32:794–805, 2019. 2
- [43] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1357–1366, 2017. 2
- [44] Jason J Yu, Adam W Harley, and Konstantinos G Derpanis. Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness. In *Eur. Conf. Comput. Vis.*, pages 3–10, 2016. 2, 4
- [45] Feihu Zhang, Oliver J Woodford, Victor Adrian Prisacariu, and Philip HS Torr. Separable flow: Learning motion cost volumes for optical flow estimation. In *Int. Conf. Comput. Vis.*, pages 10807–10817, 2021. 2
- [46] He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 695–704, 2018. 2
- [47] Shengyu Zhao, Yilun Sheng, Yue Dong, Eric I Chang, Yan Xu, et al. Maskflownet: Asymmetric feature matching with learnable occlusion mask. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 6278–6287, 2020. 2
- [48] Tinghui Zhou, Matthew Brown, Noah Snavely, and David G Lowe. Unsupervised learning of depth and ego-motion from video. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1851–1858, 2017. 4
- [49] Yuliang Zou, Zelun Luo, and Jia-Bin Huang. Df-net: Unsupervised joint learning of depth and flow using cross-task consistency. In *Eur. Conf. Comput. Vis.*, pages 1–18. Springer, 2018. 4