Robust Unsupervised Optical Flow Under Low-Visibility Conditions

Libo Long, Tianran Liu, Robert Laganière, and Jochen Lang

University of Ottawa
{llong014,tliu157,laganier,jlang}@uottawa.ca

Abstract. Although state-of-the-art unsupervised optical flow methods achieve impressive results in clean scenes, they still struggle under low-visibility weather and illumination. However, generalizing to those conditions is essential in real-world safety-critical settings, such as autonomous driving. We tackle this issue with RobFlow: an effective unsupervised learning framework that works reliably under both, standard and low-visibility conditions. Our approach is based on observations regarding the challenges of low-visibility images. Low quality features in low-visibility weather or illumination will degrade the estimation of motion boundaries in optical flow. To address these issues, we design a local-structure recovery to bridge the difference between clean and low-quality features. Additionally, we propose a confidence map to correct possible inconsistency during unsupervised learning, which will filter out large error pixels in a comparative manner. Extensive experiments demonstrate the effectiveness of our techniques. Our method outperforms prior works in both standard and low-visibility conditions, without any increase in parameter count.

1 Introduction

Optical flow, which aims at estimating pixel-level motion for videos, is fundamental for high-level computer vision applications, such as autonomous driving, video editing or object tracking. While supervised optical flow estimation methods have achieved remarkable progress, they rely on ground-truth labels, which are expensive to collect [5, 22]. Obtaining ground-truth labels requires 3D sensors and manual effort [5, 21]. To evade these issues, geometrical constraints have been widely explored for learning optimal flow in self-supervised frameworks [8, 12, 16, 17, 38]. Such methods are inexpensive and trained with only image sequences captured from RGB cameras.

Unsupervised approaches mainly rely on brightness constancy [24, 36]. Previous methods are able to produce sharp and accurate flow in clean scenes e.g., in good ligthing conditions. Adverse weather and low-illumination conditions (e.g., fog and night) further introduce noise that affects pixel correspondence as shown in Fig. 1. However, optical flow estimation in safety-critical settings such as autonomous driving [4, 27] require robust optical flow in all conditions.

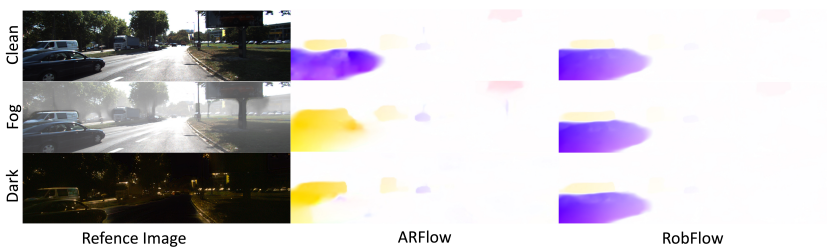

Fig. 1: ARFlow performs well on clean scenes but suffers from degradation in adverse conditions. Our method exhibits robustness to various environmental changes.

A few pioneering works have investigated optical flow estimation in adverse weather conditions. However, they are mostly tailored to specific weather scenarios and rely on highly complex pipelines and architectures [13, 35, 37, 40]. Moreover, weather-agnostic models suffer from a significant degradation when applied to a standard daytime environment [39] or need to introduce additional sensors [10, 11]. Therefore, there is a need to introduce a straightforward framework that can achieve robust optical flow estimation in both clean and low-visibility conditions with only RGB image input.

Motion boundary errors is the main error in optical flow estimation under low-visibility conditions that degrade the feature quality and blur the motion boundaries in optical flow, as shown in Fig. 1 col 2. Ideally, optical flow should be invariant for the same image pair taken under different and low-visibility conditions. In this work, we leverage the clean optical flow prediction as a reference to investigate the causes of the performance degradation under low-visibility conditions.

Specifically, we observe a clear difference in features between clean and low-visibility images. Intuitively, if the feature is invariant to weather conditions, a model will produce a consistent flow for clean and adverse scenes. Thus, we propose a local-structure recovery method that tries to minimize the structural differences between clean and degraded features by using local cosine similarity. Our key assumption is that minimizing the feature structure differences can result in more similar optical flows. Furthermore, we propose a duo-photometric loss L_{dph}, which uses the same training signal to guide both clean and noisy scenes and a Confidence Map (CM), which evaluates the reliability of each pixel in clean feature. CM is used to selectively guide the degraded feature with trustworthy pixels. Fig. 1, column 3, shows that our RobFlow method can handle various environmental changes and produces consistent flow estimates for different visibility conditions, which indicates its applicability for safety-critical applications.

We summarize the main contribution as follows:

- We propose a local-structure recovery technique to cope with motion boundary errors caused by low-visibility images such as during fog and at night time. We propose a reliable self-guidance method that filters out unreliable pixels during self-supervised learning and increases stability.

- Our proposed method simultaneously improves optical flow estimation in clean scenes and in low-visibility conditions for synthetic and real-world data, without requiring additional parameters.

2 Related Work

Optical Flow Estimation. Optical flow estimation is the task of estimating pixel-level motion between video frames. Recently, many deep learning based methods have been proposed to learn optical flow in a supervised manner [7, 9, 18, 29–31, 34]. PWC-Net [30] proposed a warp, cost volume architecture that estimates optical flow in a coarse-to-fine pyramid. RAFT [31] improved it by replacing the encoder with GRU and proposed a 4D cost volume that matches all pairs of pixels. GMA [9] further proposed a transformer module to capture hidden motion. To deal with the lack of real data, unsupervised methods [8, 15, 17, 20, 28] have been proposed with photometric loss. Selflow [17] proposed to distill reliable flow estimations from non-occluded pixels, and use these predictions as pseudo-label to learn optical flow for hallucinated occlusions. ARFlow [15] proposed to use the original image as a signal to guide augmented images.

Optical Flow under Challenging condition. Optical flow estimation has achieved remarkable results in clean scenes [7–9, 15, 17, 20, 30, 31]. However, estimation remains challenging in the presence of adverse weather or illumination effects, such as rain, fog, or low-light. Previous research has mainly focused on addressing specific weather effects by designing tailored methods or datasets. RobustFlow [13] was the first method to estimate optical flow in rainy scenes, by using a handcrafted residue channel and its color variant as a prior that is invariant to rain streaks. RainFlow [14] proposed a feature mapping operation that automatically learned rain-invariant and veiling-invariant features in a supervised manner. Zheng et al. [37] developed a method to synthesize large-scale low-light optical flow datasets by simulating the noise model on dark raw images. Yan et al. [35] proposed DenseFogFlow, a semi-supervised method that modeled the flow consistency between clean images and their corresponding rendered foggy images, to overcome the difficulty of collecting ground truth for real fog scenes. Schmalfuss et al [26] propose that augmenting the training data with weather effects enhances the robustness of optical flow methods in cross-dataset evaluation. Zhou et al. [40] presented an unsupervised domain adaptation method that adapted the model from synthetic fog images to real fog images. They propose to address motion boundary error by transferring the knowledge from a clean weather network to an adverse weather network. In contrast, we propose to address this issue by directly recovering the clean features from the adverse weather features in a single network, which also preserves the performance for clean images. GyroFlow [10] resorted to a hardware scheme to handle adverse weather, and utilized gyroscope data to obtain ego-motion labels of the camera for weakly-supervised optical flow estimation. GyroFlow+ [11] extended GyroFlow by proposing a self-guided fusion module that fused the background gyro field with the optical flow to obtain more detailed motions. However, this

approach was still limited by the hardware and the availability of ego-motion labels. Compared with previous methods, our method is the first to handle both clean and low-visibility conditions in an unsupervised manner, without requiring any additional sensors or labels.

3 Method

3.1 Overview

In this paper, we introduce RobFlow, a self-supervised architecture that robustly estimates optical flow in both clean and low-visibility (e.g., fog, night-time). Our approach utilizes the effectiveness of previous methods in clean scenes to reduce the motion errors in adverse settings.

We highlight the trade-off existing models face in clean and adverse conditions and the lack of generalization to unseen data. The pipeline of our novel framework is shown in Fig. 2. In our experiments, we use SelFlow [17] and ARFlow [15] as baseline methods. Our method does not increase parameters and memory during testing.

3.2 Preliminaries

Clean-to-adverse translation. To achieve the above, we need correspondence between low-visibility and clean samples. Following [25], we use a physics-based renderer (PBR) [32] to generate fog. CoMoGan [23] is used to generate night time scenes for the KITTI dataset. These two methods can preserve the original image as the background, and overlay adverse effects in the foreground.

We apply the above methods to generate one-to-many projections, where one clean image corresponds to fog and night images. During training, we choose a clean and a random adverse image as a corresponding pair. For clarity, we denote $I_c = (I_c^{(1)}, I_c^{(2)})$ as clean images, and denote $I_a = (I_a^{(1)}, I_a^{(2)})$ as corresponding degraded images because of, e.g., fog. We select ARFlow [15] as backbone. In each training step, the network predicts **clean flow** $\Omega(I_c)$ by $(I_c^{(1)}, I_c^{(2)})$ and generates **degraded flow** $\Omega(I_a)$ by $(I_a^{(1)}, I_a^{(2)})$. Following the original methods, we keep the smoothing loss along with the respective method's specific constraints for the clean flow $\Omega(I_c)$, and we propose our loss function as extra constraints to enhance the network's robustness for all the above conditions.

3.3 Local-Structure Recovery

Duo-Photometric Loss. In unsupervised optical flow estimation, a photometric loss L_{ph} is used as the main constraint. Low-visibility weather breaks the brightness constancy [39], previous methods [39, 40] use clean optical flow as a pseudo-label to guide degraded flow with a self-supervised loss, as shown by L_{self} in Eq. (1).

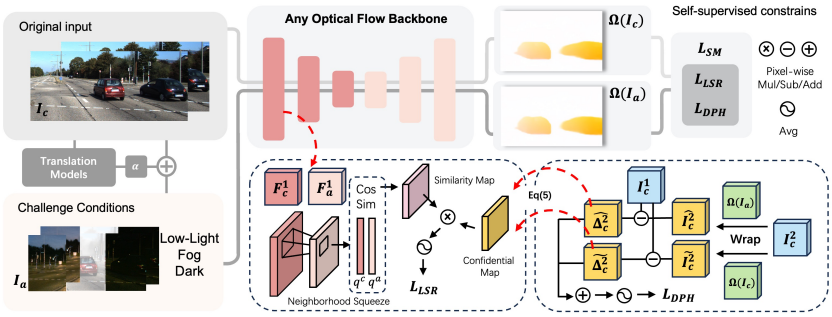

Fig. 2: Overall architecture of RobFlow. The corresponding low-visibility pairs are generated before training. In each step, we feed the original and a random low-visibility pair to the network, and align the feature and motion by our proposed methods.

$$L_{ph} = \sum |I_c^{(1)} - W(I_c^{(2)} \Omega(I_c))| \quad L_{self} = |\Omega(I_c) - \Omega(I_a)| \tag{1}$$

However, this approach is constrained by the accuracy of the clean flow estimation, as the clean flow may generate errors during unsupervised training, which may mislead the network when guiding the degraded flow. Moreover, based on Eq. (1), the clean and degraded flows are trained with different signals: the estimation of clean flow is learned by the photometric loss while estimation of degraded flow is learned from the clean flow. This strategy could result in a discrepancy between the final outputs of clean and adverse conditions.

As an example, imagine in the i-th iteration, we generate a clean flow from a clean image and a fog flow from a corresponding fog image. We use the clean flow to guide the fog flow, which will produce errors if any pixels of the clean image have lower accuracy than the fog flow. Therefore, we propose an efficient strategy that lets the clean flow and the fog flow learn from the same signal, i.e., the clean image. By exploiting the corresponding pairs of clean and adverse images, we can leverage the clean image as guidance in calculating the photometric loss for the fog optical flow. The duo-photometric (DPH) loss L_{dph}, can be defined as follows.

$$L_{dph} = \sum |I_c^{(1)} - W(I_c^{(2)}, \Omega(I_c)| + \sum |I_c^{(1)} - W(I_c^{(2)}, \Omega(I_a))| \tag{2}$$

To reveal the impact of different adverse conditions on the feature extraction of unsupervised models, we conduct analytical experiments on the synthetic KITTI [5] dataset.

Taking ARFlow as an example, we compared the pixel-wise difference of features from the encoder between clean and foggy images. As illustrated in the left panel of Fig. 3, there is a clean contrast inside the red rectangle between the clean and fog features extracted by ARFlow. As a result, the optical flow for foggy images exhibit a blurred motion boundary in the same region compared to the clean flow. This leads us to hypothesize that having similar feature maps in the encoder part will produce more consistent optical flows.

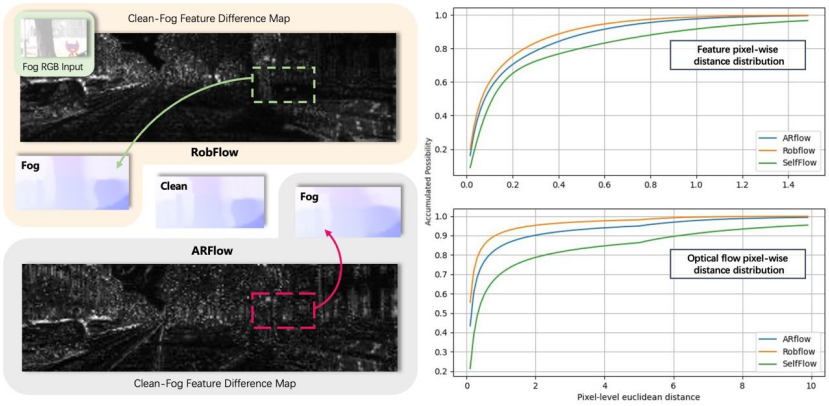

Fig. 3: The relation between features and optical flow. On the left, we show the features difference map between clean and foggy images. Our RobFlow method generates more similar features for clean and foggy images, and the optical flow in foggy conditions is closer to the optical flow in clean conditions. We report the Cumulative Distribution Function (CDF) of the distance between clean and foggy conditions in the KITTI dataset [5]. The top right figure shows the pixel-wise distance between clean and fog features. The bottom right figure shows the pixel-wise accuracy of fog optical flow and ground truth.

Methods	Clean		Fog		Dark	
	EPE	F1-all	EPE	F1-all	EPE	F1-all
ARFlow† [15]	<u>2.84</u>	<u>9.87</u>	5.43	15.79	8.22	26.85
ARFlow [15]	4.51	14.42	4.78	14.17	7.23	23.58
ARFlow-D	3.27	10.91	4.27	13.28	6.82	20.75
Selflow [17]	5.58	15.83	5.27	15.64	8.13	26.61
Selflow-D	5.02	15.12	4.92	14.83	8.54	27.25
GMA [9]	3.32	10.73	<u>3.60</u>	<u>12.39</u>	5.14	<u>17.90</u>
UPFlow-D [19]	2.81	10.03	4.32	14.58	<u>4.55</u>	18.23
RobFlow	**2.68**	**9.36**	**2.79**	**9.87**	**3.47**	**12.96**

Table 1: Quantitative results on KITTI datasets: D stand for the denoise module, we use AECR-Net [33] for defog and MCR [2] for dark images, †: only training on clean dataset, zeroshot inference on low-visibility conditions. **Bold** text represents the best result for the metric for each test. Whereas, the <u>underlined</u> text represents the second best for the metric of each test.

Therefore, we aim to minimize the discrepancy between the clean and low-visibility weather features. However, the features from different domains are difficult to project in a homogeneous space [39]. Instead of forcing them to have the same features, we relax the objective to have similar local structures for motion estimation. To this end, we propose the local-structure recovery (LSR) loss. Given the input I_c and I_a, the features $\mathcal{F}_c^i$ and $\mathcal{F}_a^i$ from the i-th block of the encoder can be obtained. Using $\mathcal{N}(p_{ij}, d)$ to represent the d-order neighbourhood of pixel p_{ij} in the feature space, $\bigcup$ represent the flatten operation, the local information $\mathbf{q}$ can be defined as

$$\mathbf{q} = \bigcup_{\mathcal{N}(p_{ij},d)} p_{ij}. \tag{3}$$

Next, we try to recover $\mathbf{q}^a$ from the corresponding $\mathbf{q}^c$. According to our observation in Sec. 3.3, the flow from clean features is not guaranteed to be correct. The clean features may also not guaranteed to be better than degraded feature for each pixels. Intuitively, we could reduce the error by ignoring those unreliable pixels. Thanks to our DPH Eq. (2), both optical flow estimations are optimized using the same clean reference image. We can obtain two error maps by computing the difference between the reference image and the target image warped by either the clean or the degraded flow. By comparing the corresponding pixels in the two error maps, the confidence map is determined by whether a pixel in the clean flow has smaller errors than the degraded flow. Inspired by [6], we use it to filter out unreliable pixels if the clean flow error is larger than the degraded flow error. The remaining reliable pixels p_{ij} are obtained by filtering with $\mathcal{K}$ are defined as follows:

$$\mathcal{K} = \left\{ p_{ij} \mid |I_c^{(1)} - \mathcal{W}(\Omega\left(I_c\right), I_c^{(2)})| < |I_c^{(1)} - \mathcal{W}(\Omega\left(I_a\right), I_c^{(2)})|, p_{ij} \in I_c^{(1)} \right\} \tag{4}$$

Finally, we use the cosine similarity to align features and the local-structure recovery loss can be described by

$$\mathcal{L}_{lsr} = \frac{1}{|\mathcal{K}|} \sum_{p_{ij} \in \mathcal{K}} 1 - \frac{\mathbf{q}^a \acute{\mathbf{q}}^c}{\|\mathbf{q}^a\| \, \|\acute{\mathbf{q}}^c\|} \tag{5}$$

$\|\|$ denotes the Euclidean norm and $\acute{}$ denotes the stop gradients. In practice, we use the intermediate flow of the same level as the feature map to generate the confidence map for the feature with the same resolution. Fig. 3 top left, visualizes the difference map of foggy and clean features of RobFlow. Our method effectively reduces the distance between the clean and foggy features, and predicts a more accurateflow in the green rectangle. Fig. 3 top right shows the distribution of the level-wise feature distance between fog and clean condition for the KITTI dataset. We observe that our RobFlow has a statistically smaller distance than ARFlow and Selflow. The bottom right figure shows the pixel-wise accuracy of fog optical flow and ground truth (clean flow is closer to the GT overall). By considering these two figures, we observe that the similarity of fog features and

clean features is positively correlated with the similarity of fog optical flow and clean optical flow which supports our claim.

The total loss of our RobFlow framework is described as follows:

$$\mathcal{L}_{all} = L_{dph} + \omega_1 L_{lsr} + \omega_2 L_{sm} + \omega_3 L_{aug} \tag{6}$$

where L_{sm} is smooth loss [20] and L_{aug} is augmentation loss [15].

Methods	L_{dph}	L_{self}	LSR	C.M	Clean		Fog		Dark	
					EPE	F1-all	EPE	F1-all	EPE	F1-all
ARFlow† [15]					2.84	9.87	5.43	15.79	8.22	30.04
ARFlow [15]		✓			11.72	28.04	12.19	30.66	14.02	26.85
ARFlow [15]		✓⋆			4.51	14.42	4.78	14.17	7.23	23.58
RobFlow	✓				3.21	12.24	3.94	13.94	5.62	18.76
RobFlow	✓		✓		3.18	12.35	3.67	12.26	4.27	16.59
RobFlow	✓		✓	✓	2.91	10.71	3.42	11.68	3.92	15.27
RobFlow	✓	✓⋆	✓	✓	**2.68**	**9.36**	**2.79**	**9.87**	**3.47**	**12.96**

Table 2: Ablation study: L_{dph} indicates Eq. (2), L_{self} indicates Eq. (1), LSR indicate the local-similarity recovery without confidence map, C.M indicates confidence map. ✓⋆ is a small scale loss (e.g. 0.01).

4 Experiments

4.1 Experiments Setup

Datasets. We conduct experiments on synthetic and real-world datasets. Our synthetic dataset is based on KITTI 2015 [5], we generate fog with physics-based rendering [32], and generate night time images by CoMoGAN [23]. We use EPE [3] and F1-all [5] metrics for evaluation.

Comparison Methods. We select recent SOTA unsupervised methods SelFlow [17], ARFlow [15], and UPFlow [19] which are designed for clean scenarios and select SOTA supervised methods GMA [9], (recent Fog optical flow methods [35] and [40] are not public code or weight). For a fair comparison, supervised methods were pre-trained on synthetic datasets with optical flow labels and then fine-tuned on target datasets with a self-supervised strategy [28]. Unsupervised methods were directly trained on the KITTI multiple-view dataset without seeing testing data. We also experiment with extra training strategies for unsupervised methods by first denoising images (deraining or defogging) during training before testing in the denoised images. Our RobFlow uses ARFlow [15] as the backbone. RobFlow* uses Selflow [17] as the backbone. For a fair comparison, we adopt the same hyperparameter settings from the original works,

such as batch size, learning rate, optimizer, and image augmentation. Neither of the proposed methods introduces any additional parameters compared to the original method.

4.2 Experiments on KITTI

Tab. 1 compares the results of our model, RobFlow, with previous state-of-the-art methods on the Synthetic KITTI 2015 test set. We make the following observations: Both RobFlow and RobFlow* outperform the original network on both clean and low-visibility conditions. RobFlow outperforms ARFlow on the original KITTI dataset, and achieves clearly better results than all previous state-of-the-art methods on the adverse weather test set. As a baseline model, ARFlow [15] (Row 1), trained only with clean KITTI images, fails to generalize to adverse weather conditions. ARFlow [15] (Row 2), trained with both clean and weather images, improves its performance on the adverse weather test set, but degrades its performance on the clean test set. The denoising methods (rows 3 & 5) improve the results for foggy images, but worsen the results for night images, because they unintentionally damage to the image details negating the benefits of noise removal.

Fig. 4 demonstrates three samples from the KITTI dataset. We find that all methods perform well on the original KITTI images. However, ARFlow [15] and UPFlow [19] both fail under foggy or dark conditions with the same scenes. In (a), ARFlow breaks the car into two pieces in low-visibility conditions, while UPFlow generates noisy motions and incorrect shapes of the car. In (b) and (c), ARFlow tends to ignore parts of objects, while UPFlow merges objects together, which could provide incorrect information in an autonomous driving system. However, our RobFlow maintains consistent results across all conditions, demonstrating its stability.

Methods	Snow		I.C.		Complex	
	EPE	F1-all	EPE	F1-all	EPE	F1-all
SelfFlow [17]	5.82	16.03	8.41	27.24	9.87	51.27
ARFlow [15]	4.89	14.86	7.72	23.71	8.05	48.12
GMA [9]	3.54	11.04	5.64	18.14	5.97	23.41
RobFlow	**2.81**	**10.42**	**2.96**	**10.50**	**3.52**	**13.21**

Table 3: Quantitative results on Zero-shot experiments.

4.3 Ablation study.

This section presents an ablation study to examine the contribution of each component in our method. We evaluate our model on both clean and adverse weather datasets to demonstrate its robustness and generalization in Tab. 2.

The first row shows the results of ARFlow trained with only clean images, which lacks the ability to handle adverse conditions.

Self-supervision (rows 2 & 3). When training the network with Eq. (1) that uses clean optical flow to guide adverse condition optical flow, the clean flow may have errors that affect the backpropagation and degrades the model performance for all images. When we reduce the self-supervision loss, the network improves the results for adverse condition images, but deteriorates the results for clean images. This trade-off exists in most of the previous methods.

Duo-photometric loss (row 4). We optimize the network for both the clean and adverse conditions data simultaneously using the duo-photometric loss. We observe improvements for both clean images and adverse condition images. However, the performance of clean scenes is still not comparable with the original network (row 1), which motivates us to further reduce motion errors.

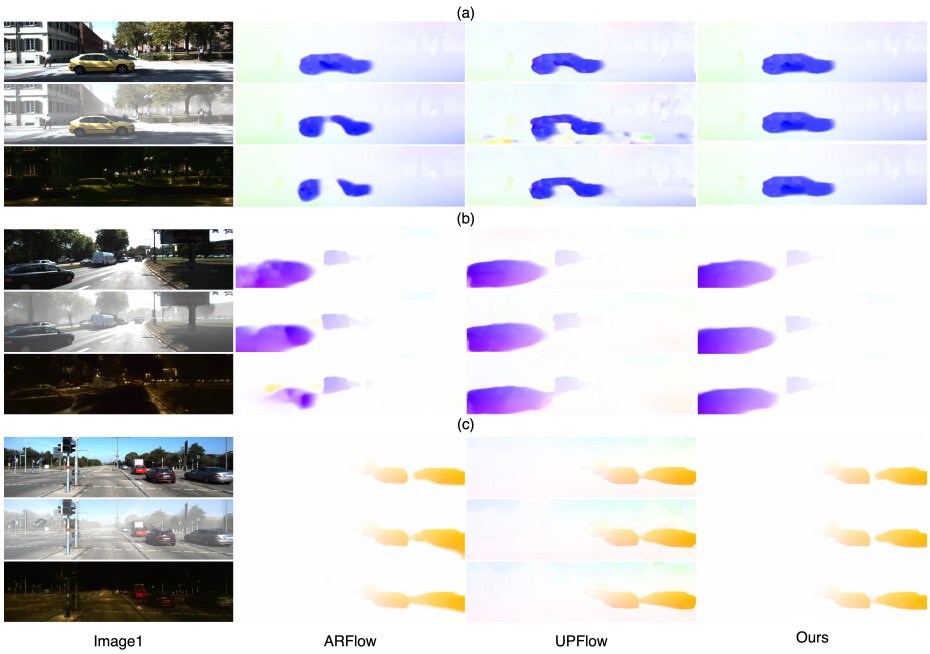

Fig. 4: Qualitative results of optical flows on KITTI.

Local-structure recovery (rows 5 & 6). We investigate the effects of LSR (Eq. (6)) and the confidence Map (Eq. (4)). First, we reduce the difference between clean and weather features by exploiting the consistency of clean images. This leads to a significant improvement in both clean and images in adverse conditions, especially in dark images. Second, we evaluate the impact of the confidence map that filters out unreliable pixels for guidance. This also results in

a reduction of the loss for all test sets. This confirms our hypothesis that reducing the errors in guidance benefits both clean and adverse conditions performance.

ALL (row 7). Our model trained with all of the contributions leads to clean results that surpass ARflow and have the best bad weather performance in row 8. We show that all components are effective in reducing the metrics for adverse conditions while maintaining the performance of clean data.

4.4 Zero-shot Experiment

To test the generalization ability of our method in real-world applications, we also conduct zero-shot experiments on different domains that are not seen during training. We use CoMoGAN [23] to synthesize images with different lighting conditions, such as dawn and dusk, and [1] to generate images with snow from the KITTI 2015 dataset. Then, we combine these techniques to create more complex scenarios, such as fog+dusk, fog+night. We use the same weights that have been trained on previous experiments and evaluate the respective method on this new test set without any fine-tuning. Tab. 3 reports the results of the zero-shot experiments in five categories: snow, illumination change (I. C.) (dawn and dusk) and complex (fog+dusk, fog+night, etc) for synthetic data. Our method outperforms all comparison and shows robustness on unseen data, which indicates that it could be a potential method for real-world applications. We collect more complex real-world data and evaluate in the supplemental material.

Index	Fog	Dark
L1	1.36	2.41
Cosine Similarity	1.20	2.08
LSR(1)	**0.94**	**1.87**
LSR(2)	1.02	1.91

Table 4: Discussion on feature loss. (i) is i-order neighbourhood.

Image Translation	Fog	Dark
Paste	1.04	-
CycleGAN [41]	-	1.94
CoMoGAN [23]	-	**1.87**
PBR [32]	**0.94**	-

Table 5: Discussion on different image translation strategies.

4.5 Discussion

Image Translation. Our proposed method requires corresponding pairs of images to establish a relationship between the clean and low-visibility conditions domains. Therefore, the choice of image translation model is crucial. We compare different strategies to generate low-visibility images. 1. We use software tools such as Adobe After Effects and Photoshop to create fog effects and overlay them on clean images. 2. We use generative adversarial networks (GANs) such as CycleGAN [41], or CoMoGAN [23] to produce realistic fog images. 3. We use physics-based rendering (PBR) [32] to simulate the optical effects of fog conditions. We training our methods whith syntetic data by the above methods,

and test in real data [11]. We present the results in Tab. 5, and observe that CoMoGAN achieves the best performance for dark images, and PBR for fog images. This is because CoMoGAN and PBR preserve the structure of the original images better than the other methods.

Importance of Duo-photometric loss. As shown in Tab. 2, our duo-photometric (DPH) loss outperforms the self-supervised loss based on clean flow. We hypothesize that the main reason is that clean flow introduces errors during training, which degrade the learning quality of flow in adverse conditions. By using the DPH loss, the flow in low-visibility conditions is less affected by the errors in the clean flow, and thus flow results are improved. Previous experiments have supported our hypothesis. We have analyzed and demonstrated that the clean flow generates errors while the degraded flow is more accurate in some pixels.

Importance of Local-structure Recovery. To evaluate the impact of the flow encoder, we conducted an experiment with three different methods for aligning clean and noisy features: L1 distance, cosine similarity, and our proposed local structure recovery. We observed that our local-structure recovery achieves better results than the other two methods. The reason is that the L1 distance or vanilla cosine similarity attempts to align pixel-level features for inputs of different domains, which is hard to achieve in a shared encoder. In an optical flow estimation model, we are more concerned about structure information (shape, geometry) than appearance information. Therefore, we only want to align local-structure information as a soft guidance.

5 Conclusion

In this work, we propose RobFlow, a simple and efficient unsupervised framework that can handle both clean and low-visibility conditions, which are a major obstacle to real-world applications (e.g., autonomous driving). To achieve this, we observed that low-quality features caused by low-visibility scenes could lead to boundary errors. Thus, we use local-similarity recovery to reduce the distance between features of clean and low-visibility images. We also propose a duo-photometric loss to reduce errors in unsupervised training. We have performed comprehensive experiments to demonstrate the robustness and generalization ability of the proposed RobFlow.

References

1. Automold–road-augmentation-library public. https://github.com/UjjwalSaxena/Automold–Road-Augmentation-Library, accessed: 2024-02-26 11
2. Dong, X., Xu, W., Miao, Z., Ma, L., Zhang, C., Yang, J., Jin, Z., Teoh, A.B.J., Shen, J.: Abandoning the bayer-filter to see in the dark. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17431–17440 (2022) 6
3. Dosovitskiy, A., Fischer, P., Ilg, E., Häusser, P., Hazırbaş, C., Golkov, V., v.d. Smagt, P., Cremers, D., Brox, T.: FlowNet: Learning optical flow with convolutional networks. In: IEEE/CVF International Conference on Computer Vision (ICCV). pp. 2758–2766. Santiago, Chile (2015) 8
4. Feng, Y., Zhang, R., Du, J., Chen, Q., Fan, R.: Freespace optical flow modeling for automated driving. IEEE/ASME Transactions on Mechatronics pp. 1–10 (2023) 1
5. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3354–3361. Providence, RI, USA (June 2012) 1, 5, 6, 8
6. Godard, C., Mac Aodha, O., Firman, M., Brostow, G.J.: Digging into self-supervised monocular depth estimation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3828–3838 (2019) 7
7. Hur, J., Roth, S.: Iterative residual refinement for joint optical flow and occlusion estimation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). p. 4761–4770. Long Beach, CA, USA (2019) 3
8. Janai, J., G"uney, F., Ranjan, A., Black, M.J., Geiger, A.: Unsupervised learning of multi-frame optical flow with occlusions. In: European Conference on Computer Vision (ECCV). vol. Lecture Notes in Computer Science, vol 11220, p. 713–731. Springer, Cham, Munich, Germany (September 2018) 1, 3
9. Jiang, S., Campbell, D., Lu, Y., Li, H., Hartley, R.: Learning to estimate hidden motions with global motion aggregation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9772–9781 (2021) 3, 6, 8, 9
10. Li, H., Luo, K., Liu, S.: Gyroflow: Gyroscope-guided unsupervised optical flow learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 288–304 (2021) 2, 3
11. Li, H., Luo, K., Zeng, B., Liu, S.: Gyroflow+: Gyroscope-guided unsupervised deep homography and optical flow learning. International Journal of Computer Vision **132**(1), 1–18 (2024) 2, 3, 12
12. Li, J., Zhao, J., Song, S., Feng, T.: Occlusion aware unsupervised learning of optical flow from video. In: Thirteenth International Conference on Machine Vision. vol. 11605, p. 224 – 231. International Society for Optics and Photonics, SPIE, Rome, Italy (2021) 1
13. Li, R., Tan, R.T., Cheong, L.F.: Robust optical flow in rainy scenes. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 288–304 (2018) 2, 3
14. Li, R., Tan, R.T., Cheong, L.F., Aviles-Rivero, A.I., Fan, Q., Schönlieb, C.B.: Rainflow: Optical flow under rain streaks and rain veiling effect. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2440–2449 (2019) 3
15. 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: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6488–6497. Virtual (June 2020) 3, 4, 6, 8, 9
16. Liu, P., King, I., Lyu, M.R., Xu, J.: Ddflow: Learning optical flow with unlabeled data distillation. In: AAAI Conference on Artificial Intelligence. p. 8571–8578. Honolulu, HI, USA (2019) 1
17. Liu, P., R.Lyu, M., King, I., Xu, J.: Selflow: Self-supervised learning of optical flow. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). p. 6707–6716. Long Beach, Ca, USA (2019) 1, 3, 4, 6, 8, 9
18. Lu, Y., Wang, Q., Ma, S., Geng, T., Chen, Y.V., Chen, H., Liu, D.: Transflow: Transformer as flow learner. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18063–18072 (2023) 3
19. Luo, K., Wang, C., Liu, S., Fan, H., Wang, J., Sun, J.: Upflow: Upsampling pyramid for unsupervised optical flow learning. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1045–1054. Virtual (June 2021) 6, 8, 9
20. Meister, S., Hur, J., Roth, S.: UnFlow: Unsupervised learning of optical flow with a bidirectional census loss. In: AAAI Conference on Artificial Intelligence. p. 7255–7263. New Orleans, La, USA (February 2018) 3, 8
21. Menze, M., Heipke, C., Geiger, A.: Joint 3d estimation of vehicles and scene flow. In: ISPRS Workshop on Image Sequence Analysis (ISA) (2015) 1
22. Menze, M., Heipke, C., Geiger, A.: Object scene flow. ISPRS Journal of Photogrammetry and Remote Sensing (JPRS) (2018) 1
23. Pizzati, F., Cerri, P., de Charette, R.: Comogan: continuous model-guided image-to-image translation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14288–14298 (2021) 4, 8, 11
24. Ren, Z., Yan, J., Ni, B., Liu, B., Yang, X., Zha, H.: Unsupervised deep learning for optical flow estimation. In: AAAI Conference on Artificial Intelligence. vol. 31 (2017) 1
25. Saunders, K., Vogiatzis, G., Manso, L.: Self-supervised monocular depth estimation: Let's talk about the weather. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1234–1243 (2023) 4
26. Schmalfuss, J., Mehl, L., Bruhn, A.: Distracting downpour: Adversarial weather attacks for motion estimation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 10106–10116 (October 2023) 3
27. Shi, H., Zhou, Y., Yang, K., Yin, X., Wang, K.: Csflow: Learning optical flow via cross strip correlation for autonomous driving. In: 2022 IEEE Intelligent Vehicles Symposium (IV). pp. 1851–1858 (2022) 1
28. Stone, A., Maurer, D., Ayvaci, A., Angelova, A., Jonschkowski, R.: SMURF: Self-teaching multi-frame unsupervised raft with full-image warping. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3887–3896. Nashville, TN, USA (June 2021) 3, 8
29. Sui, X., Li, S., Geng, X., Wu, Y., Xu, X., Liu, Y., Goh, R., Zhu, H.: Craft: Cross-attentional flow transformer for robust optical flow. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 123–132 (2022) 3
30. Sun, D., Yang, X., Liu, M.Y., Kautz, J.: PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). p. 8934–8943. Salt Lake City, UT, USA (2018) 3

31. Teed, Z., Deng, J.: RAFT: Recurrent All-Pairs Field Transforms for Optical Flow. In: European Conference on Computer Vision (ECCV). p. 402–419. Glasgow, UK (2020) 3
32. Tremblay, M., Halder, S.S., De Charette, R., Lalonde, J.F.: Rain rendering for evaluating and improving robustness to bad weather. International Journal of Computer Vision **129**(2), 341–360 (2021) 4, 8, 11
33. Wu, H., Qu, Y., Lin, S., Zhou, J., Qiao, R., Zhang, Z., Xie, Y., Ma, L.: Contrastive learning for compact single image dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10551–10560 (June 2021) 6
34. Xu, H., Zhang, J., Cai, J., Rezatofighi, H., Tao, D.: Gmflow: Learning optical flow via global matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 795–804 (2022) 3
35. 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 (CVPR). pp. 288–304 (2020) 2, 3, 8
36. Yu, J.J., Harley, A.W., Derpanis, K.G.: Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness. In: Computer Vision–ECCV 2016 Workshops. pp. 3–10. Amsterdam, The Netherlands (October 2016) 1
37. Zhang, M., Zheng, Y., Lu, F.: Optical flow in the dark. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(12), 9464–9476 (2022) 2, 3
38. Zhong, Y., Ji, P., Wang, J., Dai, Y., Li, H.: Unsupervised deep epipolar flow for stationary or dynamic scenes. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). p. 12069–12078. Long Beach, Ca, USA (2019) 1
39. Zhou, H., Chang, Y., Chen, G., Yan, L.: Unsupervised hierarchical domain adaptation for adverse weather optical flow. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 3778–3786 (2023) 2, 4, 7
40. Zhou, H., Chang, Y., Yan, W., Yan, L.: Unsupervised cumulative domain adaptation for foggy scene optical flow. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 288–304 (2023) 2, 3, 4, 8
41. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networkss. In: Computer Vision (ICCV), 2017 IEEE International Conference on (2017) 11