CEDFlow: Latent Contour Enhancement for Dark Optical Flow Estimation

Fengyuan Zuo[1], **Zhaolin Xiao**[1, 2*], **Haiyan Jin**[1, 2], **Haonan Su**[1, 2]

[1]Xi'an University of Technology, China, 710048
[2]Shaanxi Key Laboratory for Network Computing and Security Technology, China, 710048
xiaozhaolin@xaut.edu.cn, jinhaiyan@xaut.edu.cn

Abstract

Accurately computing optical flow in low-contrast and noisy dark images is challenging, especially when contour information is degraded or difficult to extract. This paper proposes CEDFlow, a latent space contour enhancement for estimating optical flow in dark environments. By leveraging spatial frequency feature decomposition, CEDFlow effectively encodes local and global motion features. Importantly, we introduce the 2nd-order Gaussian difference operation to select salient contour features in the latent space precisely. It is specifically designed for large-scale contour components essential in dark optical flow estimation. Experimental results on the FCDN and VBOF datasets demonstrate that CEDFlow outperforms state-of-the-art methods in terms of the EPE index and produces more accurate and robust flow estimation. Our code is available at: https://github.com/xautstuzfy.

Introduction

Optical flow estimation is a crucial technique of numerous computer vision applications, such as autonomous driving (Takumi et al. 2017), object tracking (Peng et al. 2020), video enhancement (She and Xu 2022), *etc*.. Under a global scene smoothness assumption, researchers propose to estimate the optical flow by solving a global energy minimization problem (Horn and Schunck 1981). If assuming the key points have brightness constancy, the flow estimation can also be formulated as a local energy minimization (Lucas, Kanade et al. 1981). However, in dark illumination scenarios, these assumptions can be violated due to low contrast, strong noise, and deterioration of brightness constancy. As shown in Fig. 1, the contour information in low-contrast dark images is degraded by intense noise, leading to ambiguous contour matching. This presents a significant challenge in achieving precise flow estimation in such conditions.

Pre-stage image feature enhancement has emerged as a promising approach to address the challenge of Dark Optical Flow Estimation (DOFE). While deep learning-based solutions have made remarkable progress in enhancing low-light or dark images (Li et al. 2021), existing methods primarily focus on improving visual perceptual quality by adjusting brightness and contrast. Nevertheless, these enhancements often introduce inconsistencies and blurry boundaries, providing limited benefits to specific vision tasks. In contrast, task-specific feature-level enhancement has shown effectiveness in applications like face detection (Wang et al. 2022), image deblurring (Zhou, Li, and Change Loy 2022), and image or video super-resolution (Chan et al. 2021). This paper proposes a novel feature enhancement framework explicitly designed for DOFE, distinguishing it from existing low-light image enhancements.

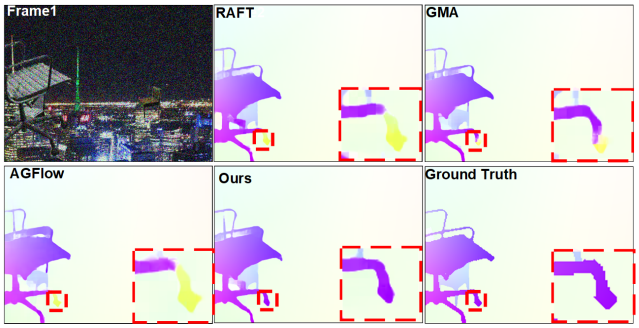

Figure 1: Flow estimation in a challenging low-light condition. The proposed CEDFlow outperforms state-of-the-art methods RAFT (Teed and Deng 2020), GMA (Jiang et al. 2021a) and AGFlow(Luo et al. 2022b).

Large-scale background motion poses challenges for DOFE, where both local and global features play crucial roles. Global feature extraction usually requires a larger receptive field size, which is computationally expensive. Meanwhile, local feature extraction is scale-sensitive, especially in the presence of low-light noises. Therefore, choosing an appropriate receptive field size that enables the simultaneous extraction of local and global motion features is difficult. Furthermore, large-scale salient contour semantics are very important for precise DOFE, but accurately picking the salient contour semantics from a low-light image is also challenging. To address these issues, we propose CEDFlow, an efficient latent contour encoding and enhancement in DOFE. Our contributions can be summarized as follows.

- **A spatial frequency decomposition for local and global motion encoding.** We propose encoding frequency-based features through local and global

*Corresponding author:Zhaolin Xiao, Haiyan Jin

motion encoders, which can be integrated after feature attention has been computed by MLP.

- **A latent space contour enhancement.** We suggest computing the 2nd-order Gaussian difference of the feature map to select large-scale contour semantics. This process enables the direct enhancement of contour features in the latent space while accentuating local discrimination and smoothness.
- **The state-of-the-art performance on widely used benchmarks.** The proposed CEDFlow outperforms state-of-the-art approaches regarding the End-Point Error index on the public FCDN and VBOF benchmarks.

Previous Works on the DOFE

To address optical flow estimation with small moving objects and occlusions, convolutional neural networks (CNNs) have been successfully applied in methods like FlowNet (Dosovitskiy et al. 2015), Pyramid-networks (Sun et al. 2018), RAFT (Teed and Deng 2020), GMA (Jiang et al. 2021a), and GMFlow (Xu et al. 2022a). However, these state-of-the-art methods heavily rely on high-contrast image textures, which can be significantly degraded in DOFE.

A straightforward solution for DOFE is to enhance low-light input images using computational enhancements, which are dominated by learning-based solutions, like (Guo, Li, and Ling 2016; Cai et al. 2017; Wei et al. 2018; Wang et al. 2020). Recent solutions effectively improve the visual perceptual quality of low-light images by using frequency adaptive operations (Xu et al. 2020), (Xu et al. 2022b). However, none of these works are specifically designed for the DOFE problem. Aiming at solving the DOFE problem, Zheng et al. propose a synthetic optical flow benchmark by adding dark image noise to the FlyChairs dataset, called FlyingChairs Dark&Noise (FCDN) dataset (Zheng, Zhang, and Lu 2020). They also introduce the Various Brightness Optical Flow (VBOF) dataset, which includes multiple exposure levels and optical flow pseudo labels (Zhang, Zheng, and Lu 2021).

Few works currently focus on designing specific DOFE-oriented learning networks, which is more helpful than applying general-purpose low-light enhancements. Therefore, our proposed CEDFlow framework explores enhancing salient contour semantics, which is essential for large-scale motion understanding, specifically addressing the DOFE.

The Proposed CEDFlow Algorithm

Fig. 2 illustrates the decomposition of consecutive frames into high- and low-frequency parts, then enabling the extraction of fine-grained and large-scale motion information through local and global encoders. Furthermore, we suggest computing the 2nd-order Gaussian difference of the latent feature map to select and to enhance the salient contour semantics.

Motion Feature Encoding

Different from mainstream motion encoders (Teed and Deng 2020; Luo et al. 2022c; Xu et al. 2022a) as shown in Fig. 3(a), we suggest encoding high- and low-frequency components with local and global encoders after spatial frequency decomposition, *i.e.*Fig. 3(b). The challenge of long-range pixel connections in DOFE arises from noise hindering pixel matching under low-light conditions. While increasing the receptive field with larger convolution kernels can be a solution, it may introduce pixel similarity uncertainty and feature-matching ambiguity. Instead, we propose a context-adaptive motion reasoning approach to construct long-term and short-term pixel correlations. The motion encoder in our method consists of a spatial frequency decomposition, a dual-branch motion encoder(DBME), and a Multilayer Perceptron (MLP)-based feature aggregation.

The Frequency-based Decomposition. To begin with, we introduce a spatial frequency decomposition module that first utilizes three downsampling blocks to extract a feature map volume $f(H/8\times W/8\times N)$ from the input frames, where H and W represent the height and width of the frames, respectively, and N denotes the number of channels. This downsampling step helps reduce the computational cost and compress the motion representation. To decompose the motion feature information, we use two groups of dilated convolutions (with kernel size/dilation rate of 1/1 and 3/2), denoted as d_1 and d_2. By computing the convolutional difference between d_1 and d_2, a contrast-aware attention map ω_s can be defined as,

$$\omega_s = sigmoid(d_1(f) - d_2(f)), \tag{1}$$

With the weight map ω_s, the extracted feature volume f can be roughly divided as the high-frequency part f^H and the low-frequency part f^L.

$$< f^L, f^H >=< (1-\omega_s)\cdot f, \omega_s \cdot f >, \tag{2}$$

where "$\cdot$" denotes an element-wise dot-product. An example feature map visualization can be found in Fig. 4. After the feature decomposition, the f^L represents the spatial low-frequency properties, and the f^H concerns the high-frequency feature of the given dark scene.

The Dual-branch Motion Feature Encoder. Instead of extracting multiple-scale features, we introduce a dual-branch structure to encode motion features. Our feature extraction consists of two key components: the global and local encoders. The decomposed high-frequency part f^H contains high-contrast shape and regional motion, as shown in Fig.4, but often mixed with strong dark noise. To effectively capture local and structural motion while mitigating additional noise interference, the local encoder is equipped with three 2D convolutional blocks (with a small receptive field). For the low-frequency part f^L, we perform large-kernel 1-D convolutional blocks (with a large receptive field) to learn X-direction and Y-direction motions, followed by layer normalization to ensure robust propagation of motion information. The architecture of the local encoder is depicted in Fig. 2. The learned feature contains local and structural motion features, which are utilized to make more accurate predictions for the DOFE problem. Inspired by constructing the channel attention (Hu, Shen, and Sun 2018), we suggest an MLP-based aggregation module to bridge the gap between

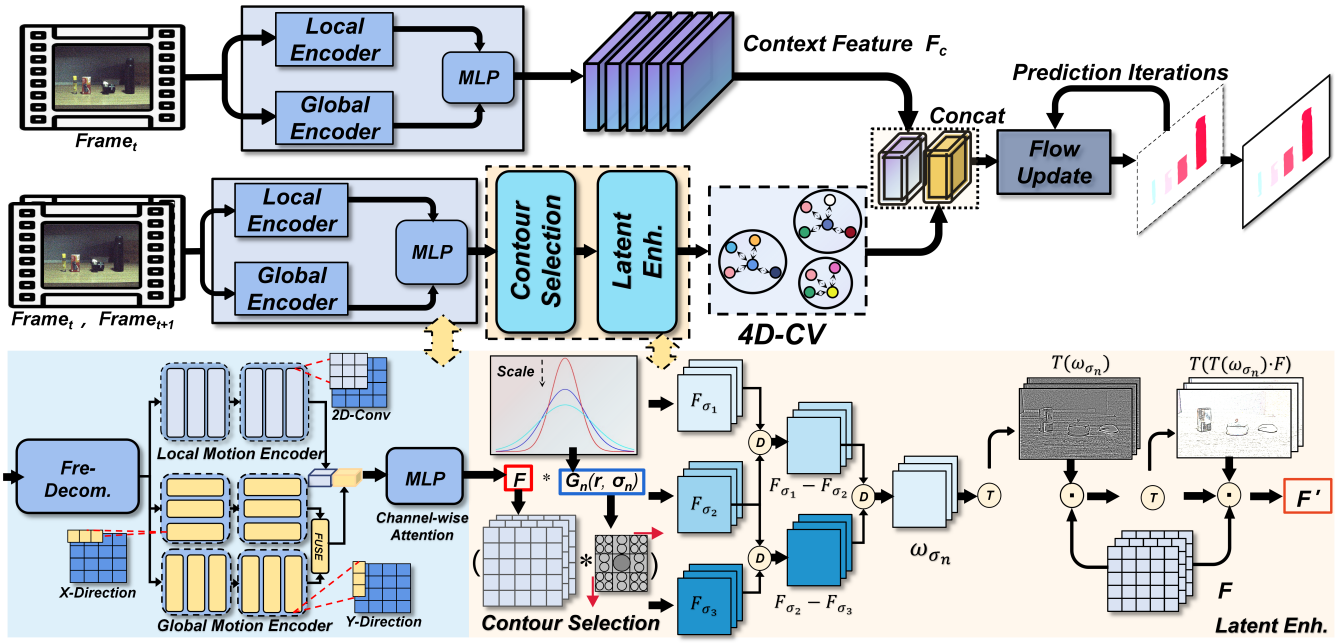

Figure 2: The proposed latent contour enhancement architecture. It mainly consists of a feature decomposition-based motion encoder and a latent contour enhancement module. The core of our contour enhancement adopts the D^{2}oG operation that directly selects the large-scale contours in the latent space. "D" is a subtraction operation; "T" denotes the sigmoid function; "·" represents the dot product; "F" means a motion-encoded feature map; "F'" is a contour-enhanced feature map.

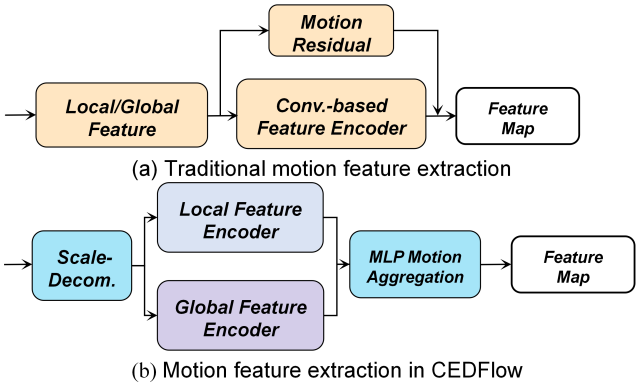

Figure 3: Highlight the feature extractions of the proposed CEDFlow. Most methods extract motion features using a single feature encoder. In contrast, CEDFlow utilizes distinctively structured encoders to separately encode local and global motion, which are aggregated using an MLP.

local and global motion representation, which can be formulated as:

$$F = MLP(\mathcal{G}(f^L), \mathcal{L}(f^H)), \tag{3}$$

where $MLP(\cdot)$ denotes the aggregation module, F is the aggregated feature map, the $\mathcal{G}(\cdot)$ and $\mathcal{L}(\cdot)$ indicate the global encoder and the local encoder. In general, $\mathcal{G}(f^L)$ and $\mathcal{L}(f^H)$ are used to extract the long-range and short-range connections from low- and high-frequency parts, respectively. Then, an $MLP(\cdot)$ is applied for the feature aggregation.

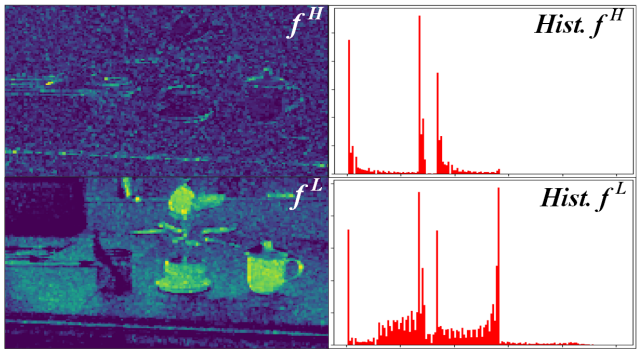

Figure 4: A feature map visualization of frequency-based components. Histograms show different feature distributions between f^L and f^H. The high-intensity pixels in f^H concentrate on areas with structural significance, *i.e.*, shape and region with motion.

Latent Contour Enhanced Flow Estimation

To improve the motion reasoning of the DOFE, we propose a novel approach for latent contour selection and enhancement. Unlike traditional methods operating in the spatial or frequency domain, our approach focuses on contour selection and strengthening in the latent space directly. Rather than trivial contours, we specifically target large-scale contours, which are more critical to estimation reliability.

Large-scale Latent Contour Selection. In the DOFE computation, large-scale contour plays an important role in con-

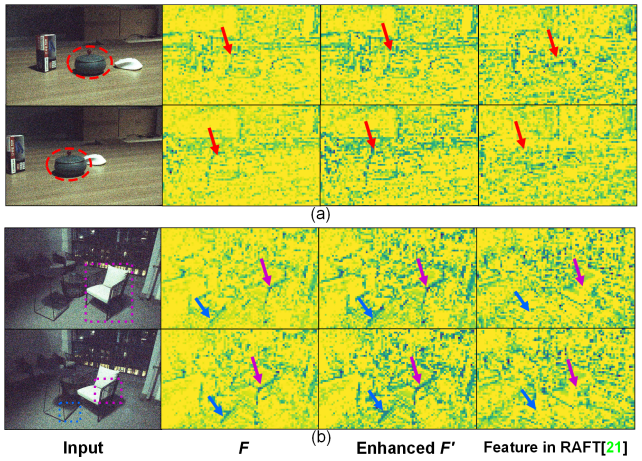

Figure 5: Two sets of input frame pairs and their corresponding feature maps are depicted in (a) and (b) respectively. The enhanced feature map F' exhibits improved preservation of object contours compared to RAFT's(Teed and Deng 2020) corresponding layered feature map. Moreover, the saliency of large-scale contours is enhanced in F', obtained from the feature map F using the D^{2}oG contour selection method.

straining the motion correlation to consecutive image areas with the same motion. We propose using pre-defined Gaussian kernels to compute the difference between the extracted feature vector and its neighborhood and to select these salient contour semantics. While Gaussian-like difference computation is commonly considered practical in the image spatial domain, we explore using this computation directly on the feature embedding maps, where the network can learn the latent features. Specifically, we apply Gaussian blurring to the feature map F with different standard deviations σ_n and radius r to obtain the Gaussian-blurred feature maps F_{σ_n} as follows:

$$F_{\sigma_n} = F * G_n(r, \sigma_n), \quad n = 1, 2, 3. \tag{4}$$

Instead of using the 1st-order Gaussian difference, we propose to use the 2nd-order Gaussian difference of the feature map as a weighting function,

$$\omega_\sigma = F_{\sigma_1} - 2F_{\sigma_2} + F_{\sigma_3}, \tag{5}$$

the ω_σ represents the saliency of feature vectors in F, where a large ω_σ value indicates that the corresponding feature refers to a more salient large-scale contour semantics. By setting different σ_n and r, we can determine the scale and saliency of contour selection. Empirically, a large r and significant difference between σ_n values can filter out more trivial contours, leaving only large-scale and high-contrast contours. The suggested parameter settings for the FCDN and VBOF datasets are $r = 3.0$, resulting in a Gaussian kernel size of 7×7, and $\sigma_1 = 3, \sigma_2 = 9, \sigma_3 = 27$ respectively. A detailed comparison of different parameter settings can also be found in the experiments section. In summary, by using Gaussian blurring with different standard deviations and radii, we compute the 2nd-order Gaussian difference of

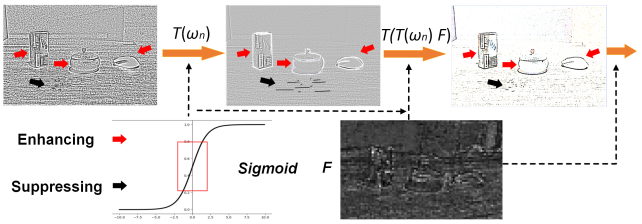

Figure 6: The proposed double-pass filtering contour enhancement in the latent space.

the feature map to enhance the saliency of large-scale contours, called D^{2}oG contour selection, which offers a robust and accurate optional process for incorporating flow estimation. Fig. 5 displays the encoded and aggregated motion feature map F. By observing this figure, it can be seen that the motion features have almost the same intensity. This implies that the weight or saliency of these features cannot be distinguished when computing the optical flow using the feature map F. However, after contour enhancement with suitable computation using F and ω_σ in latent space, the large-scale contours become more visible than other features in the F'.

Latent Enhancement and Flow Estimation. After the large-scale latent contour selection, we present a latent space enhancement that enables a more precise DOFE. The double-pass filtering is applied on the salient contour weighting ω_σ to highlight the large-scale contours and alleviate the trivial high-contrast feature. Instead of directly multiplying ω_σ to F, the double-pass filtering is performed in the latent space with sigmoid normalization. This process is shown in Fig. 6, where the enhanced large-scale contours are more visible while the trivial high-contrast contours are reduced. Eq. 4 demonstrates how this filtering is applied to the latent feature map F.

$$F' = T(T(\omega_\sigma) \cdot F) \cdot F, \tag{6}$$

where $T(\cdot)$ stands for the sigmoid normalization. Fig. 5 visually illustrates the proposed large-scale contour enhancement. We suggest using double-pass filtering to enhance large-scale contours (pointing by the red arrows) and suppress trivial contours (pointing by the black arrows). Since the proposed $T(\cdot)$ transformation-based contour enhancement directly operates the aggregated feature map F in the latent space, our computational cost is significantly lower than other spatial or frequency image processing operations. The proposed latent space operations enable end-to-end learning-based DOFE training and prediction.

To establish feature correspondences, we adopt the approach from previous successful work (Teed and Deng 2020) and compute the 4D correlation volume (4D-CV), representing the pixel-wise correspondence between two feature maps of the input paired frames. The following equation constructs visual similarity between all pairs of feature vectors in the two contour-enhanced features F'_1 and F'_2,

$$CV = correlation(F'_1, F'_2). \tag{7}$$

By pooling the last two dimensions of the original correlation volume, both large and small displacements of pixel

correlation can be better encoded and searched using a pyramid structural multiple-layered 4D-CV.

The flow estimation module in CEDFlow is based on the flow update module of RAFT (Teed and Deng 2020). A GRU-based update operator is employed to iteratively update the flow estimation result by looking up values from the 4D-CV. We initialize the flow field p^0 to zero, and in the k-th iteration, it produces an update flow p^k_Δ, which is added to the current estimate: $p^k = p^k_\Delta + p^{k-1}$. To compute the update flow p^k_Δ, we utilize the current flow estimate p^k to retrieve correlation features from the correlation pyramid 4D-CV,

$$\begin{aligned} p^k_\Delta &= GRU(p^{k-1}, F_c) \\ &= (1-\rho)\cdot p^{k-1} + \rho \cdot \phi(p^{k-1}, F_c). \end{aligned} \tag{8}$$

In equation (8), $\phi(\cdot)$ is a $Tanh()$-based activation of the current flow update increment, which jointly considers the context feature F_c, as shown in Fig. 2 and the last flow estimation p^{k-1}. The parameter ρ is an automatically computed weighting factor that balances the update state and the reset state of the GRU flow estimation. Here, ρ is calculated by sequentially applying concatenation, convolution, and sigmoid activation on the p^{k-1} and F_c,

$$\rho = Sigmoid \circ Conv \circ Concat(p^{k-1}, F_c). \tag{9}$$

A more detailed explanation of the GRU computation can be found in (Teed and Deng 2020) or in our code.

To train the proposed model in a supervised manner, we employ a simple $\mathcal{L}-1$ loss to constraint the differences between the predicted optical flow p^k and corresponding ground truth p_{gt}:

$$\mathcal{L} = \sum_{k=1}^{K} \gamma^{K-k} \left\| p^k - p_{gt} \right\|_1 . \tag{10}$$

In our experiments, we set $\gamma = 0.9$ cooperating with many flow prediction iterations ($K = 12$), enabling a better coarse-to-fine flow updating.

Experiments

Analysis on Different Parameters Settings

Since latent D^{2}oG operation is sensitive to the settings of the Gaussian kernel parameters, we first studied the different setting combinations of the parameters σ_n and radius r. As shown in Tab. 1, on the FCDN and VBOF (Fuji2) datasets, we first analyzed the performance of the CEDFlow by switching r when using a fixed $\sigma_n = \{3, 9, 27\}$. The best choice of the radius is $r = 3.0$, *i.e.*, using a 7×7 kernel size, our CEDFlow achieves the mean values of EPE metric are 1.08 and 13.94 on the FCDN and VBOF datasets respectively. Further, in the σ_n analysis, we found that more extensive settings of σ_n accompanied by a larger receptive field lead to better performance in constructing long-range pixel correlation. However, increasing the perceptive field is computationally expensive, *e.g.* , when using a Gaussian kernel with $r = 3.0$, the network in CEDFlow consists of approximately 7.7 million parameters. However, if we increase the kernel size to $r = 4.0$, corresponding to a 9×9 kernel, the parameter count rises to 8.7 million, nearly 1. million the number of parameters compared to $r = 3.0$. Additionally, when setting $r = 4.0$, each training iteration takes approximately 7.4% more time than $r = 3.0$.

Parameter	EPE(Trained FCDN)		
	Settings	FCDN	VBOF(Fuji2)
r	1	1.27	14.22
	2	1.17	13.99
	3	**1.08**	**13.94**
	4	1.21	14.15
σ_n	(1, 3, 9)	1.17	14.14
	(2, 6, 18)	1.18	14.33
	(3, 9, 27)	**1.08**	**13.94**

Table 1: EPE comparison of different parameters setting.

Comparison with State-of-the-Arts

We evaluated the CEDFlow against eight state-of-the-art methods that have achieved top-performing results on the Sintel (Butler et al. 2012) and KITTI (Menze, Heipke, and Geiger 2015) leaderboards. These methods include RAFT (Teed and Deng 2020), GMFlow (Xu et al. 2022a), GMFlowNet (Zhao et al. 2022), GMA (Jiang et al. 2021a), AGFlow (Luo et al. 2022c), KPAFlow (Luo et al. 2022a), SCV (Jiang et al. 2021b), and Flow1D (Xu et al. 2021). All models are fairly trained on the FCDN and Mix (FCDN + VBOF) datasets. Please see details of the experiment implementation in the supplementary.

Training on the FCDN Dataset. Tab. 2 presents the evaluation results of nine models on FCDN and VBOF (Fuji2 part only). Our CEDFlow achieved the best average end-point error (EPE) of 1.08 on the FCDN (the 2nd column). The CEDFlow outperforms the second-ranked AGFlow (Luo et al. 2022c) about 7% in EPE (1.15 $\rightarrow$ 1.08), and outperforms GMFlowNet (Zhao et al. 2022) near to 30.7% in EPE (1.56 $\rightarrow$ 1.08). These results indicate that CEDFlow outperforms other models in solving the DOFE problem, thanks to the support of feature decomposition-based motion learning and latent space contour enhancement.

In Tab. 2 (3rd and 4th columns), our proposed CEDFlow achieved the best performance on VBOF (Fuji2) and VBOF (All) datasets, outperforming eight state-of-the-art approaches. It improved by 3.2%, 4.3%, and 5.2% over GMA, GMFlow, and KPAFlow on Fuji2, respectively. AGFlow obtained the second-best scores, and RAFT and GMA ranked third. CEDFlow demonstrated excellent cross-data capabilities and superior performance.

Training on the Mixed Datasets. In Tab. 2 (5th, 6th and 7th columns), we trained all models using Mixed (FCDN + VBOF) datasets, then the flow estimation evaluations are presented on the FCDN, Fuji2 and VBOF datasets respectively. The proposed CEDFlow achieved an EPE index of 1.23 on the FCDN, 4.69 on the Fuji2 and 6.52 on the VBOF, outperforming the other eight models. On the FCDN dataset, GMA (Jiang et al. 2021a) won second place with an

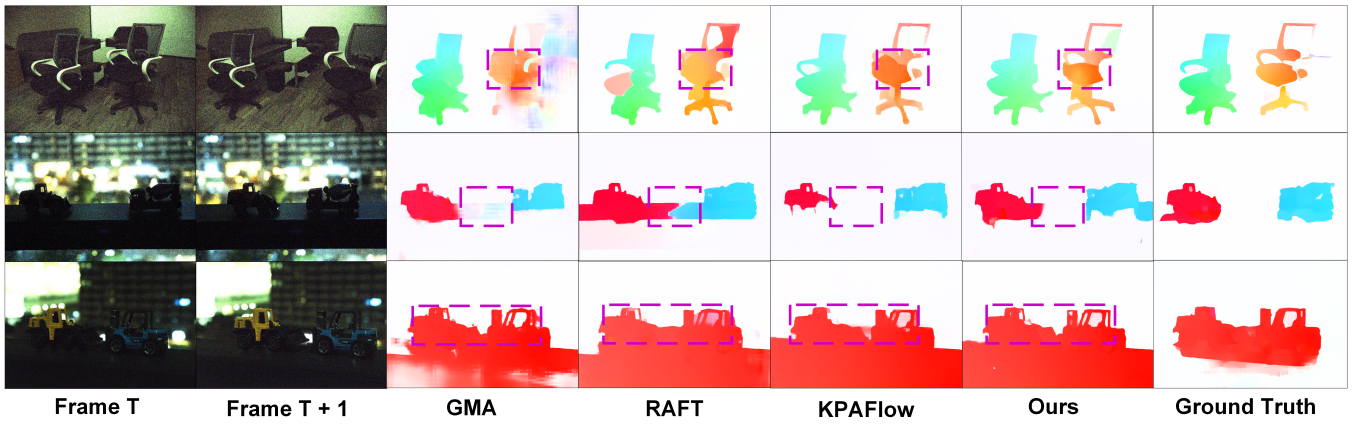

Figure 7: Visual comparison of different flow estimations. In strong noise conditions, our CEDFlow outperforms state-of-the-art methods in terms of precision (1st row). Furthermore, the 2nd and 3rd rows demonstrate CEDFlow's distinct and accurate contour structures, closely resembling the ground truth (highlighted in boxes).

EPE	Trained on FCDN			Trained on Mixed		
	FCDN	VBOF(Fuji2)	VBOF(All)	FCDN	VBOF(Fuji2)	VBOF(All)
RAFT(Teed and Deng 2020)	1.23	14.20	21.84	1.38	7.34	8.89
GMFlowNet(Zhao et al. 2022)	1.56	14.51	22.87	1.70	7.71	8.66
GMFlow(Xu et al. 2022a)	1.18	14.56	22.72	1.31	5.76	7.23
AGFlow(Luo et al. 2022c)	<u>1.15</u>	<u>14.16</u>	<u>21.05</u>	1.27	4.97	<u>6.75</u>
GMA (Jiang et al. 2021a)	1.18	14.40	21.77	<u>1.26</u>	<u>4.90</u>	6.81
KPAFlow(Luo et al. 2022a)	1.24	14.71	23.10	1.39	6.11	7.47
SCV (Jiang et al. 2021b)	1.29	14.96	24.13	1.27	6.48	7.76
Flow1D (Xu et al. 2021)	1.22	14.25	21.79	1.30	5.13	6.93
CEDFlow(Ours)	**1.08**	**13.94**	**20.89**	**1.23**	**4.69**	**6.52**

Table 2: EPE comparison of different flow evaluations on FCDN and VBOF datasets. Underlining denotes second rank.

EPE index of 1.26, 2.4% higher than our CEDFlow's 1.23 score. CEDFlow outperforms the GMFlowNet by nearly 28% in EPE, which indicates the effectiveness of CEDFlow in addressing the DOFE problem. On the Fuji2, GMA also achieved the second-best performance with an EPE index of 4.90, close to CEDFlow. Due to the more complex composition of the Mix dataset, models trained on Mix generally had higher EPE indexes compared to models trained on FCDN only. In the 6th column of Tab. 2, we present a comparison of the entire VBOF dataset, which includes a wide range of scenarios with illumination changes. The CEDFlow model remains the best-performing one with an EPE index of 6.52. The AGFlow approach follows closely behind with a score of 6.75. However, RAFT and GMFlowNet models achieved the lowest scores, indicating that most state-of-the-art flow estimations are less effective for significant illumination variations.

Visual Comparison. Fig. 7 visually compares our flow estimation method, CEDFlow, with RAFT, GMA and KPAFlow algorithms on the VBOF dataset (Trained on FCDN). In low-light conditions, the results of three representative scenarios demonstrate that the proposed CEDFlow provides accurate and robust DOFE. The proposed CEDFlow outperforms other algorithms in generating accurate contours as in Fig. 7 first row, it can be observed that the results obtained from GMA and RAFT are not as accurate as the proposed CEDFlow method, especially in terms of the generated contours. For instance, the chair's cushion in the middle of the image appears jagged and blurry in the results obtained from GMA and RAFT, whereas CEDFlow gener-

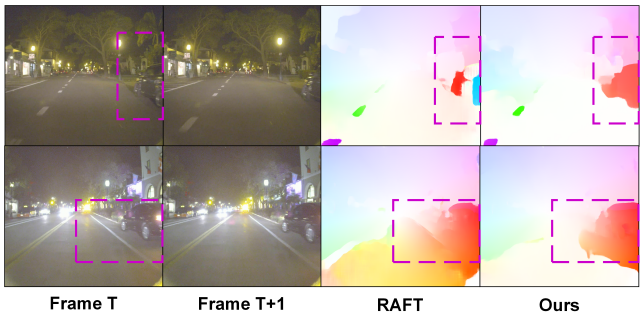

Figure 8: Visualization results on FLIR ADAS dataset. CEDFlow demonstrates superior performance in handling complex motion.

Framework	EPE(Trained on FCDN)	
	FCDN	VBOF(Fuji2)
RAFT	1.23	14.20
RAFT+DBME	1.14	14.13
RAFT+LCE	1.12	14.03
CEDFlow+RAFT Encoder	1.17	14.38
CEDFlow	1.08	13.94

Table 3: EPE comparisons by switch encoders between RAFT and CEDFlow frameworks.

ates a more precise contour closer to the ground truth. We compared DOFE visual results on the FLIR ADAS dataset (available at github), which includes many real-world driving scenes in the dark. Although the FLIR ADAS has no optical flow ground truth, we can see that CEDFlow performs better in this challenging scenario, which is with moving objects and dynamic lighting conditions, as shown in Fig. 8. In general, the proposed CEDFlow outperforms SOTAs in terms of precision under low-light conditions.

Ablation Studies

Comparison when using Different Encoders. To validate the effectiveness of our proposed Dual-Branch Motion Encoder (DBME), we conducted an ablation experiment by switching encoders in both CEDFlow and RAFT frameworks. All tested models were trained on the FCDN dataset. As shown in Tab. 3, we replaced the original encoder of the RAFT with the proposed DBME (2nd row). Furthermore, we added our Latent Contour Enhancement (LCE) module in the RAFT framework (3rd row). We have selected the parameter r to achieve optimal performance. For the 4th row, we replaced the DBME in CEDFlow with the RAFT encoder. Tab. 3 demonstrates that the proposed DBME and LCE perform best in CEDFlow and are also effective in other flow estimations, *e.g.* , the RAFT.

Ablation With CEDFlow Components. A qualitative comparison is provided in Tab. 4, presenting the analysis results of the DBME and LCE components. When removing the feature decomposition module (1st row of Tab. 4), we observe that the EPE index of CEDFlow increases by 9.6%, indicating a significant performance degradation. A larger increase in EPE signifies a more significant impact of the removed module or component on performance improvement. By removing the global or local encoder separately (2nd and 3rd rows), we demonstrate that the global encoder contributes more precision than the local encoder. The results in Tab. 4 (5th row) highlight the substantial contribution of our LCE module compared to other modules. This further emphasizes the effectiveness of the proposed latent contour enhancement in improving flow estimation performance.

Computation Analysis

Parameters. In the 2nd column of Tab. 5, we compare the parameter capacity of different SOTAs. Our CEDFlow has 7.7 million parameters, the second largest model. It is because CEDFlow employs the DBME that encodes local and global motion features separately, and the additional parameters of DBME have proved valuable for performance.

Module	EPE (Trained on FCDN)		
	FCDN	VBOF(Fuji2)	VBOF(All)
w/o Decom.	1.19	14.70	23.13
w/o Glo. Enc.	1.17	14.54	22.64
w/o Loc. Enc.	1.15	14.23	21.88
w/o MLP	1.22	15.01	24.02
w/o LCE	1.26	14.79	23.44
Whole	1.08	13.94	20.89

Table 4: Ablation analysis for different parts of the DBME and the LCE in CEDFlow framework.

Models	Param(M)	Time(ms)	Memory(GB)
RAFT	5.3	42	1.7
GMFlowNet	9.3	112	3.4
GMFlow	4.7	67	1.8
AGFlow	5.6	46	1.9
GMA	5.9	63	1.8
KPAFlow	5.8	89	2.6
SCV	5.3	40	1.6
Flow1D	5.7	45	1.7
Ours	7.7	76	2.1

Table 5: Comparisons of the EPE and computational costs with the state-of-the-art methods.

Runtime & Memory. We also show the runtime and memory requirements of different models in Tab. 5. As inputting images at 736×480 resolution, our CEDFlow requires 76ms in runtime and 2.1GB in memory. Considering the significant improvement the precision, its computational costs are acceptable for dealing with the challenging DOFE problem.

Conclusion

This paper proposes a novel CEDFlow framework for dense optical flow estimation that addresses the challenges in low-light conditions. CEDFlow incorporates the Dual-Branch Motion Encoder (DBME) and Latent Contour Enhancement (LCE) modules to improve accuracy and robustness. The DBME captures finer details by utilizing its distinctively structured local and global motion feature encoders, while the LCE module enhances large-scale contours in the latent feature space. Experimental results on FCDN and VBOF datasets demonstrate that CEDFlow outperforms state-of-the-art methods regarding end-point error. Future research directions include exploring the application of CEDFlow to other vision tasks and investigating optimizations for further enhancing efficiency and accuracy.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (62272383, 62371389, 62031023).

References

Butler, D. J.; Wulff, J.; Stanley, G. B.; and Black, M. J. 2012. A naturalistic open source movie for optical flow evaluation. In *ECCV*, 611–625. Springer.

Cai, B.; Xu, X.; Guo, K.; Jia, K.; Hu, B.; and Tao, D. 2017. A joint intrinsic-extrinsic prior model for retinex. In *ICCV*, 4000–4009.

Chan, K. C.; Wang, X.; Yu, K.; Dong, C.; and Loy, C. C. 2021. Basicvsr: The search for essential components in video super-resolution and beyond. In *CVPR*, 4947–4956.

Dosovitskiy, A.; Fischer, P.; Ilg, E.; Hausser, P.; Hazirbas, C.; Golkov, V.; Van Der Smagt, P.; Cremers, D.; and Brox, T. 2015. Flownet: Learning optical flow with convolutional networks. In *ICCV*, 2758–2766.

Guo, X.; Li, Y.; and Ling, H. 2016. LIME: Low-light image enhancement via illumination map estimation. *TIP*, 26(2): 982–993.

Horn, B. K.; and Schunck, B. G. 1981. Determining optical flow. *Artificial intelligence*, 17(1-3): 185–203.

Hu, J.; Shen, L.; and Sun, G. 2018. Squeeze-and-excitation networks. In *CVPR*, 7132–7141.

Jiang, S.; Campbell, D.; Lu, Y.; Li, H.; and Hartley, R. 2021a. Learning to estimate hidden motions with global motion aggregation. In *ICCV*, 9772–9781.

Jiang, S.; Lu, Y.; Li, H.; and Hartley, R. 2021b. Learning optical flow from a few matches. In *CVPR*, 16592–16600.

Li, C.; Guo, C.; Han, L.; Jiang, J.; Cheng, M.-M.; Gu, J.; and Loy, C. C. 2021. Low-light image and video enhancement using deep learning: A survey. *TPAMI*, 44(12): 9396–9416.

Lucas, B. D.; Kanade, T.; et al. 1981. *An iterative image registration technique with an application to stereo vision*, volume 81. Vancouver.

Luo, A.; Yang, F.; Li, X.; and Liu, S. 2022a. Learning Optical Flow With Kernel Patch Attention. In *CVPR*, 8906–8915.

Luo, A.; Yang, F.; Luo, K.; Li, X.; Fan, H.; and Liu, S. 2022b. Learning optical flow with adaptive graph reasoning. In *AAAI*, volume 36, 1890–1898.

Luo, A.; Yang, F.; Luo, K.; Li, X.; Fan, H.; and Liu, S. 2022c. Learning optical flow with adaptive graph reasoning. In *AAAI*, volume 36, 1890–1898.

Menze, M.; Heipke, C.; and Geiger, A. 2015. Joint 3d estimation of vehicles and scene flow. *ISPRS annals of the photogrammetry, remote sensing and spatial information sciences*, 2: 427.

Peng, J.; Gu, Y.; Wang, Y.; Wang, C.; Li, J.; and Huang, F. 2020. Dense scene multiple object tracking with box-plane matching. In *ACM MM*, 4615–4619.

She, D.; and Xu, K. 2022. An Image-to-video Model for Real-Time Video Enhancement. In *ACM MM*, 1837–1846.

Sun, D.; Yang, X.; Liu, M.-Y.; and Kautz, J. 2018. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In *CVPR*, 8934–8943.

Takumi, K.; Watanabe, K.; Ha, Q.; Tejero-De-Pablos, A.; Ushiku, Y.; and Harada, T. 2017. Multispectral object detection for autonomous vehicles. In *ACM MM*, 35–43.

Teed, Z.; and Deng, J. 2020. Raft: Recurrent all-pairs field transforms for optical flow. In *ECCV*, 402–419. Springer.

Wang, L.-W.; Liu, Z.-S.; Siu, W.-C.; and Lun, D. P. 2020. Lightening network for low-light image enhancement. *TIP*, 29: 7984–7996.

Wang, W.; Wang, X.; Yang, W.; and Liu, J. 2022. Unsupervised face detection in the dark. *TPAMI*, 45(1): 1250–1266.

Wei, C.; Wang, W.; Yang, W.; and Liu, J. 2018. Deep retinex decomposition for low-light enhancement. *British Machine Vision Conference*.

Xu, H.; Yang, J.; Cai, J.; Zhang, J.; and Tong, X. 2021. High-resolution optical flow from 1d attention and correlation. In *ICCV*, 10498–10507.

Xu, H.; Zhang, J.; Cai, J.; Rezatofighi, H.; and Tao, D. 2022a. GMFlow: Learning Optical Flow via Global Matching. In *CVPR*, 8121–8130.

Xu, K.; Yang, X.; Yin, B.; and Lau, R. W. 2020. Learning to restore low-light images via decomposition-and-enhancement. In *CVPR*, 2281–2290.

Xu, X.; Wang, R.; Fu, C.-W.; and Jia, J. 2022b. SNR-Aware Low-Light Image Enhancement. In *CVPR*, 17714–17724.

Zhang, M.; Zheng, Y.; and Lu, F. 2021. Optical Flow in the Dark. *TPAMI*.

Zhao, S.; Zhao, L.; Zhang, Z.; Zhou, E.; and Metaxas, D. 2022. Global Matching with Overlapping Attention for Optical Flow Estimation. In *CVPR*, 17592–17601.

Zheng, Y.; Zhang, M.; and Lu, F. 2020. Optical flow in the dark. In *CVPR*, 6749–6757.

Zhou, S.; Li, C.; and Change Loy, C. 2022. Lednet: Joint low-light enhancement and deblurring in the dark. In *ECCV*, 573–589. Springer.