ITC 1/54
Information Technology and Control
Vol. 54 / No. 1/ 2025
pp. 219-233
DOI 10.5755/j01.itc.54.1.37533

Optical Flow Estimation Method Based on Bidirectional Consistency Combined Occlusion

Received 2024/06/04 | Accepted after revision 2024/11/14

HOW TO CITE: Guo, H., Wang, Y., Guo, X. (2025). Optical Flow Estimation Method Based on Bidirectional Consistency Combined Occlusion. *Information Technology and Control*, 54(1), 219-233. https://doi.org/10.5755/j01.itc.54.1.37533

Optical Flow Estimation Method Based on Bidirectional Consistency Combined Occlusion

Haoxin Guo, Yifan Wang, Xiaobo Guo
Changchun University of Science and Technology, Changchun 130022, China

Corresponding author: gxbcust@163.com

In response to the failure of optical flow estimation to solve the tracking accuracy degradation caused by motion occlusion, this paper proposes an optical flow estimation method based on bidirectional consistency combined occlusion inference to improve the tracking accuracy degradation caused by motion occlusion. First, by utilizing the symmetry between forward and reverse optical flow mapping and occlusion mapping, the optical flow estimation value, luminance, contrast, and structure are simultaneously used as constraints for occlusion detection. Then, a new dynamic weight loss function module was designed to supervise the training of the optical flow estimation model. The endpoint error loss function is used and smooth L1 and gradient terms are introduced to obtain a continuous and smooth optical flow field, and binary cross entropy loss is used to solve the occlusion problem of consistency. Finally, experiments have shown that the proposed method outperforms FPCR Net, FlowNet3 and SCV algorithms in tracking accuracy on the MPI Sintel, Flying Chairs and KITTI datasets. Also, the proposed method achieves endpoint error values of 1.01 (Clean train), 1.07 (Final train), 0.88 (Flying Chairs), and 3.37 (KITTI) on the above datasets, respectively, and has significant advantages in resisting occlusion.

KEYWORDS: Optical flow estimation, Bidirectional structural consistency, Occlusion detection, Dynamic weight.

1. Introduction

Optical flow estimation is based on the pixel relationship between two adjacent frames in the video and searching for pixel displacement changes to determine the motion state of the target. Traditional differential-based optical flow estimation [4, 29, 38] requires multiple epochs and it is difficult to achieve both timeliness and accuracy. Deep learning-based optical flow estimation [13, 22, 1, 31, 39, 40] outperforms traditional

optical flow estimation in terms of speed and accuracy in public optical flow databases (MPI Sintel/Clean [18], FlyingChairs [5]), but also faces degradation of tracking accuracy due to motion occlusion.

Occlusion can be classified as intraclass occlusion and interclass occlusion because the measured object is hidden by the same type of object, or by fixed elements or other class objects. Occlusion is a key issue affecting the robustness of optical flow estimation [32].

Conventional optical flow estimation uses non-coherent motion for regularization to pass information from non-occluded regions to occluded regions [23]. However, when non-occluded regions are mislocalized, it leads to tracking failure. Unsupervised optical flow estimation learns the minimum photometric loss value between images from the unlabeled data. However, owing to the warp of the images, the minimum photometric loss value still cannot determine the correct optical flow in the occluded region [39, 28, 26].

In occlusion situations, when supervised optical flow estimation is trained the target may be completely or partially occluded, and differences in occlusion location and degree cause errors and inconsistencies in the label [29], doubling the difficulty of tracking. Even the advanced supervised optical flow estimation methods PWCNet [32] and FlowNet [17] face an occlusion problem. Flownet3 [15] uses interpolation to learn the occlusion region to avoid errors or missing motion compensation information between the two parallax m+aps when computing the optical flow. Sun et al. [32] jointly optimized occlusion and optical flow estimation but did not consider the intrinsic connection between the two leading to overly complex function construction, which affects the tracking speed of the final optical flow estimation. MirrorFlow [13] directly fuses the optical flow of the front and back terms in an optimization function that integrates the masking inference process in the function optimization process and uses a split-epochs method to obtain the final results. However, the excessive complexity and computational costs limit the application of this algorithm. As shown in Figure 1, the FlowNet family of methods has the best endpoint error (EPE) values in the MPI Sintel database, but its motion edges are blurred.

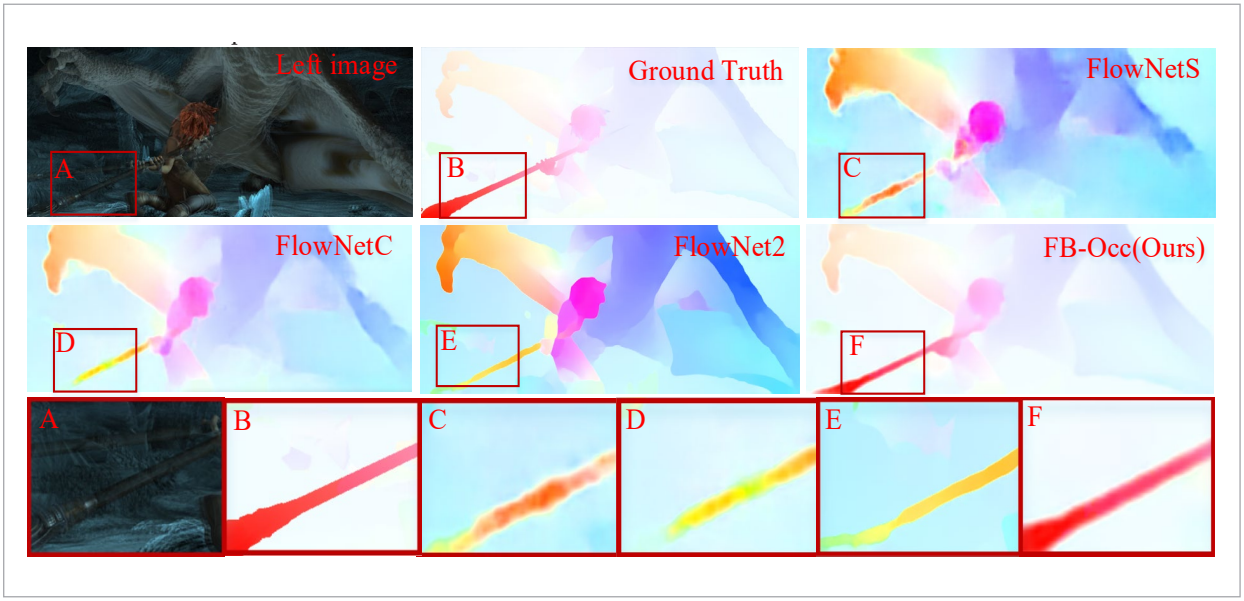
In motion occlusion scenes, the edge blurring problem caused by motion occlusion is an important challenge in optical flow estimation. The edge information of the object is blurred due to occlusion, leading to poor performance of the optical flow estimation results in the occluded regions. When dealing with motion edges, traditional methods rely on the gradient information of the image for compensation, and the gradient information can be distorted in occlusion situations, resulting in edge blurring of the optical flow estimation. While deep learning methods improve overall edge detection ability through end-to-end training, the models still struggle to accurately capture edge information in the face of complex occlusions and fast motion. In summary, when facing the occluded regions, due to the interpolation errors, optical flow computation errors, and other factors, the existing occlusion processing methods based on optical flow estimation not only increase the complexity of the algorithms, but also cause the edge blurring problem caused by motion occlusion. Therefore, how to improve the accuracy of optical flow estimation in the occluded regions while ensuring the computational efficiency is still an urgent research difficulty to be solved.

This research performs forward and backward passes by exchanging the input order of the front and back frames, and utilizes the estimation values of front and back optical flow, structure, brightness, and contrast consistency to perform occlusion inference, thereby improving the accuracy of the optical flow estimation in the occluded regions. A multi-loss function mechanism is designed to enhance the robustness of real images, avoid complex optimization process and improve the robustness of the algorithm. The proposed optical flow estimation method with bidirectional consistency joint occlusion inference in this paper has occlusion and consistency detection ability, and extracts the occluded edge information of the motion target through enhancing the consistency detection. It is called the optical flow estimation method based on bidirectional consistency combined occlusion inference, abbreviated as FB-Occ. As shown in Figure 1, the optical flow effect map of the FB-Occ method is better and has a significant edge protection advantage in the occluded regions.

The main contributions of this study are as follows:

1 A check method with optical flow estimation and bidirectional structural consistency is proposed, which combines optical flow and occlusion esti-

Figure 1
Plot of optical flow output results for the cave_4 datasets in MPI Sintel. The validation results of the left image, ground truth, FlowNetS, FlowNetC, FlowNet2, and the algorithm FB-Occ in this paper are shown, where the red box at the bottom is a close-up view



mation, and uses the occlusion information output from the encoder-decoder and SSIM as the occlusion detection result to improve the tracking accuracy in occlusion scenes.

2 By setting dynamic weights, the EPE loss, smooth L1 loss, gradient loss, and binary cross entropy loss are combined for supervised-learning optical flow estimation and occlusion estimation, reducing the effect of occlusion.

3 Experiments on the MPI Sintel, Flying Chairs and KITTI datasets show that the proposed method is superior to existing methods in resisting occlusion.

The content structure of this article is as follows. Section 1 introduces the existing problems of occlusion detection, as well as the structural arrangement and innovation points of this article. Section 2 provides an overview of the current research status of anti-occlusion optical flow estimation algorithms. In Section 3, the main techniques of the proposed method are described. Section 4 elaborates on the method presented in this article, identifying training details and evaluation indicators. Experimental verification and data analysis are conducted. Section 5 summarizes the full text and prospects.

2. Related Work

Optical flow estimation based on convolution theory is a new concept for realizing object tracking, which can achieve end-to-end learning under the condition of a large amount of data with labeled samples. Occlusion detection and optical flow estimation are two overlapping problems. This section summarizes the optical flow estimation based on convolutional neural networks and the methods used for object tracking under occlusion conditions by researchers.

2.1. Optical Flow Estimation Based on Convolutional Neural Network

FlowNet [9] is the first model to use deep learning for optical flow estimation. FlowNet2 [14] stacks basic models to improve the model capacity and performance, whereas SpyNet [30] uses image pyramids and warping to build compact models. FlowNet3 [15] achieved motion edge retention in an occlusion case through superposition and warping. PWCNet [32] used the classical optical flow principle to construct a widely used effective model. LiteFlowNet [12] is based on the FlowNet algorithm, which constructs

a lightweight network by reducing the number of convolutional layers and filter size to reduce computational costs and memory usage. LiteFlowNet2 [11] improves the accuracy of optical flow estimation based on variational methods and lightweight cascading. RAFT [34] introduced a feature refinement module with shared weights to update the retrieval of relevant optical flow fields and reduce the complexity of searching for feature information in high-resolution images. DICL [36] and SCV [17] respectively introduce a multiscale matching module, and a K-neighborhood matching method to calculate the correlation of the same feature vector in two frames, which is used to reduce the tracking error value of each pixel.

Although convolutional neural networks have made significant progress in optical flow estimation, in recent years, Vision Transformers (ViTs), as an emerging method, have demonstrated strong capabilities in multiple computer vision tasks. ViTs can better capture the global features of an image through the self-attention mechanism and perform well in tasks such as image classification [43, 35, 6], monocular depth estimation [10], and other tasks. This provides a new research direction for the further development of optical flow estimation methods.

2.2. Occlusion and Optical Flow Estimation

Occlusion detection and optical flow estimation are two overlapping problems, and the processing of occlusion problems in object tracking mainly depends on whether the optical flow estimation is based on occlusion estimation.

2.2.1. Based on Occlusion Estimation

The limitation of the image block-based algorithm [20] lies in the one-sided partitioning method. Too dense a division cannot use the neighborhood information, and too sparse a division may divide the occlusion region and non-occlusion region together, leading to the failure of the algorithm. Matching information-based methods [2, 7] have sparse matching points, and the coverage of object occlusion leads to the disappearance of some features of the target. Thus, it is easy to judge the region where no occlusion occurs as the occlusion region because of the difficulty in matching the model, resulting in false detection. The Epicflow [42] method based on optical flow divergence compensates for occluded optical flow information by interpolation, which can affect the judgment of the occluded region owing to factors such as interpolation error or optical flow calculation error. The PMC-PWC [8] proposed a parallel edge-preserving optical flow estimation with occlusion detection based on a multiscale context. SelFlow [25] used Noc-Model from the un-occluded region to learn reliable optical flow information and then used the reliable optical flow information learned by the Noc-Model to guide the OCC-Model to learn the optical flow of the occluded pixels. This method does not consider the intrinsic connection between the occlusion problem and the optical flow estimation problem, leading to the construction of a function that is too complex and affects the speed of final optical flow estimation results.

2.2.2. Based on Non-occlusion Estimation

Non-occlusion estimation allows the algorithm to disobey the consistency assumption to a certain extent, that is, forward and reverse optical flow estimates are estimated using asymmetric methods [3], followed by bidirectional consistency calibration, and finally, the insertion of the calibration values in the anomalous pixels. OccInpFlow [28] avoids the occlusion inference process and treats occlusion points as outliers when the consistency assumption is not considered. It reduces the sensitivity of the algorithm to outliers by improving the robustness of the network structure and the feature context linkage. MaskFlowNet [42] proposed an asymmetric occlusion-aware feature-matching module that learns coarse occlusion features and filters occlusion regions immediately after a feature warp. FPCR-Net [37] combines a feature pyramid with a residual reconstruction network. The pyramid warp module uses global and local multiscale correlations to form multi-level costs by aggregating features at different scales. The residual reconstruction module reconstructed more precise residual optical flow values at each stage.

In summary, the method based on occlusion estimation can reliably detect occlusion, and the judgment of occlusion area information will increase the complexity of the algorithm. At the same time, the limitations of the algorithm include insufficient density or sparsity to utilize neighborhood information, misjudgment, and other shortcomings. The method of not performing occlusion estimation is simple and fast, but its disadvantage is that it cannot accurately process the occluded parts of the image, which may lead to incorrect recognition or tracking of the position and shape of the object. This proposes an optical flow estimation method based on bidirectional consistency joint oc-

clusion inference. This method uses the symmetry of bidirectional optical flow warp and occlusion warp, the optical flow estimation, brightness, contrast, and structure are taken as the constraints of occlusion detection, and an optimization model of dynamic weight loss function is established. This method can reliably detect occlusion and improve tracking accuracy.

3. The Network Structure of the Method in this Article

A CNN-based optical flow estimation method is constructed using an encoder and decoder. The encoder uses a convolutional neural network to complete the image motion object feature extraction for the video sequences, and the decoder uses a deconvolution network to complete the optical flow computation. Its calculation speed is fast, but its accuracy is not high in motion occlusion situations. As shown in Figure 2, to overcome the occlusion problem, this study proposes an optical flow estimation method based on bidirectional consistent joint occlusion inference.

Given two adjacent RGB images {It, It + 1: P→R^3 are two consecutive frames}, the optical flow value of forward It→It+1 is $I_f = (u_f, v_f)^T$, and the optical flow value of reverse It+1→It is $I_b=(u_b, v_b)^T$. Figure 2 outlines the network architecture of the proposed method, which can be summarized into three parts: encoder-decoder, bidirectional structure consistency check based on occlusion detection, and model optimization based on dynamic weight loss function.

3.1. Encoder-decoder Construction

Figure 3 shows a diagram of the encoder-decoder network architecture, with the structure underneath the

Figure 2
Structure diagram of optical flow estimation method based on bidirectional consistency combined occlusion inference

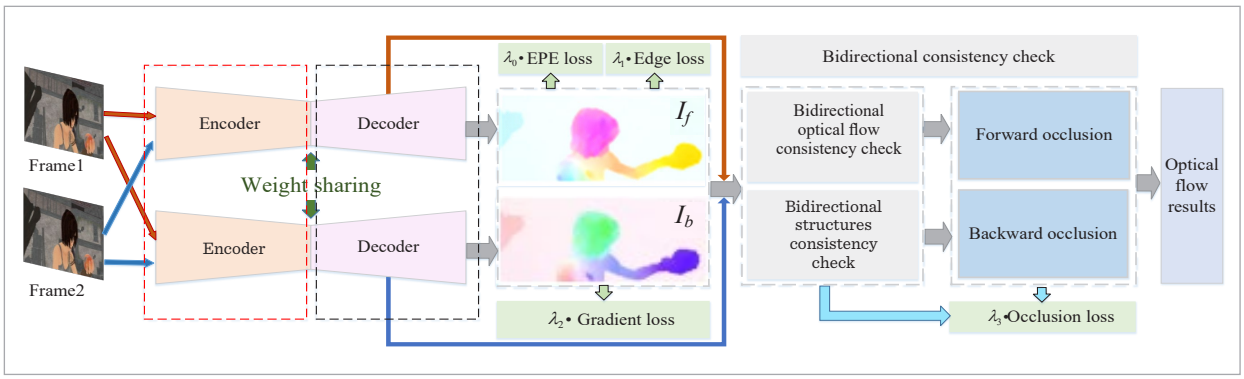
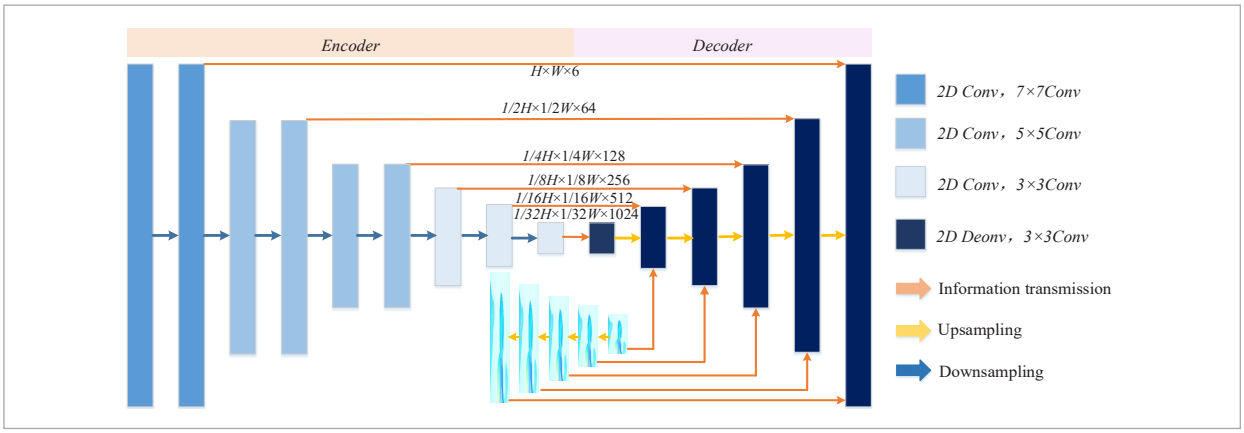


Figure 3
Structure diagram of optical flow estimation method based on bidirectional consistency combined occlusion inference



orange column on the left as the encoder architecture.

The structure below the pink column on the right side of Figure 3 is the decoder architecture, which contains 6 convolutional layers. Except for the first layer of convolution, each feature map contains three parts:

1 In a feature extraction network, a feature that matches its size.
2 Features obtained from the deconvolution of a previous small-size feature map.
3 And the features obtained from the conversion of the previous small-size feature map into a small-size optical flow field followed by deconvolution.

Finally, after concatenating the three features, the step size is 2 to obtain the next size of the input feature block, the cycle continues, and the size of the image is scaled up until it is the same size as the input image. The activation function of each layer in the entire network is LeakyReLU, allowing any image size to be input.

3.2. Bidirectional Consistent Check for Occlusion Detection

In this section, bidirectional optical flow warp and occlusion warp are estimated jointly from two adjacent RGB images, taking full advantage of the symmetry of both. Considering bidirectional consistency means that the motion of the corresponding pixel of two adjacent frames is opposite, that is, forward occlusion corresponds to reverse de-occlusion. When occlusion exists, there is no corresponding feature vector between two adjacent frames. The threshold range of the image for the normalized pixels is then adjusted to [0,1].

As shown in Figure 2, the proposed encoder-decoder is used to combine the input optical flow and structural similarity index for occlusion detection. The input of two adjacent RGB frames is passed through the encoder-decoder, and the forward and reverse optical flow values are the output. The basic idea of determining whether an occlusion is detected by optical flow bidirectional consistency.

$$\left|I_f + I_b\right|^2 < \sigma(\left|I_f\right|^2 + \left|I_b\right|^2) + \sigma_1 \cdot \tag{1}$$

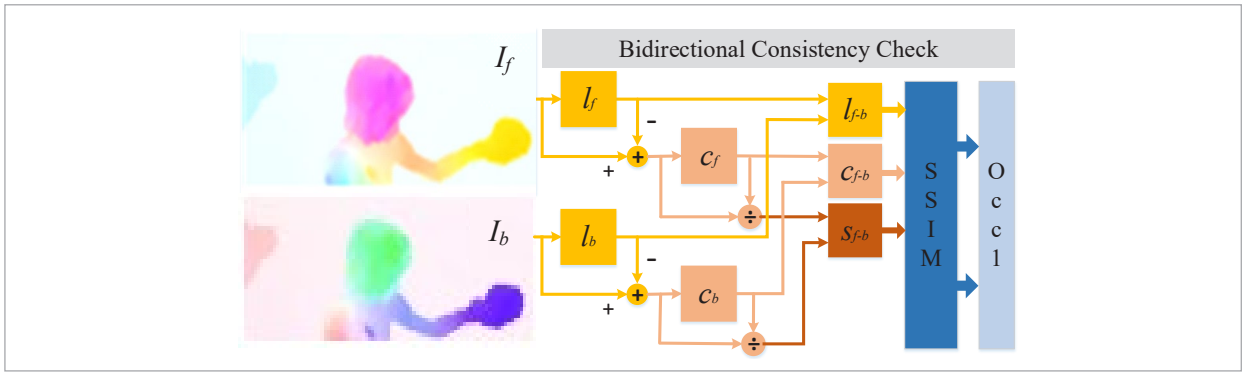
In Equation (1), I_f and I_b are the forward optical flow values and reverse optical flow values, respectively. Where σ and σ_1 are the regulation parameter: σ =0.01 and σ_1=0.5. In the absence of occlusion, the left side of Equation (1) tends to 0, marked *Occ0*=0. If Equation (1) is not satisfied, it is considered an occlusion and marked *Occ0*=1.

As shown in Figure 4, the basic idea of using the structural similarity index to determine whether there is occlusion is to compare the brightness, contrast, and structure of images in forward detection and backward detection. If the image cannot return to its original position, it is considered occlusion.

$$SSIM(I_f, I_b) = [l(I_f, I_b)]^{\alpha} \cdot [c(I_f, I_b)]^{\beta} \cdot [s(I_f, I_b)]^{\gamma} \tag{2}$$

$$\begin{aligned} l(I_f, I_b) &= \frac{2\mu_{I_f}\mu_{I_b} + C_1}{\mu_{I_f}^2 + \mu_{I_b}^2 + C_1}, \\ c(I_f, I_b) &= \frac{2\sigma_{I_f}\sigma_{I_b} + C_2}{\sigma_{I_f}^2 + \sigma_{I_b}^2 + C_2}, \\ s(I_f, I_b) &= \frac{\sigma_{I_f I_b} + C_3}{\sigma_{I_f}\sigma_{I_b} + C_3} \end{aligned} \tag{3}$$

Figure 4
Structure diagram of bidirectional consistency check



where $l(I_f, I_b)$, $c(I_f, I_b)$, and $s(I_f, I_b)$ denote the contrast of the luminance, contrast, and structure of I_f and I_b, respectively. μ_{I_f} and μ_{I_b} represent the mean values. σ_{I_f}, σ_{I_b} represents the standard deviation. $\sigma_{I_f I_b}$ is covariance. C_1, C_2, and C_3 are constants, and to ensure $l(I_f, I_b)$, $c(I_f, I_b)$, and $s(I_f, I_b)$ stability, the general parameters $\alpha = \beta = \gamma = 1$ and $C_3 = C_2/2$. It is obtained that:

$$SSIM(I_f, I_b) = \frac{(2\mu_{I_f}\mu_{I_b} + C_1)(2\mu_{I_f I_b} + C_2)}{(\mu_{I_f}^2 + \mu_{I_b}^2 + C_1)(\mu_{I_f}^2 + \mu_{I_b}^2 + C_2)}, \quad (4)$$

where $C_1 = (k_1 L)^2$, $C_2 = (k_2 L)^2$, $k_1 = 0.01$, $k_2 = 0.03$. $L = 2d - 1$ is the dynamic range of pixel values 0-255, so d=bits per pixel=8.

If two eigenvectors tend to be linearly correlated, the equation tends to be 1, otherwise it approaches 0. The smaller the SSIM value, the greater the difference between the two point eigenvectors. According to Equation (4), the unmasked marker *Occ1* =0 is obtained, and when the corresponding eigenvector is masked, the masked marker *Occ1*=1 is obtained.

As shown in Figure 2, this study's occlusion estimation network includes an encoder-decoder and a bidirectional structure consistency checking framework. Consequently, the occlusion estimation results are as follows:

$$Occ = Occ0 \,||\, Occ1 \cdot \quad (5)$$

In Equation (5), Occ0 is the result of masking the output from the encoder-decoder, and Occ1 is the masking infor-mation output from SSIM. The symbol '||' indicates the 'or' operation, and when a value of 1 exists for Occ0 or Occ1, the result is masking Occ=1; otherwise, it is 0.

3.3. Dynamic Weight Loss Function Model

In the existing research, most scholars often use the L1 function of endpoint error as the loss function of optical flow estimation, but the L1 function has poor edge detection ability. This paper combines the L1 loss function, smooth L1 loss, gradient loss function, and occlusion loss function, and these four loss functions are constrained with each other to form a new loss function to better supervise the training of the optical flow estimation model.

Most optical flow estimation methods based on convolutional neural networks use loss functions based on endpoint error (EPE) and use the L1 standard.

$$L_{flow\_epe} = \frac{1}{N}\sum_{x=0}^{W}\sum_{y=0}^{H}\left\|Flow_{out}(x,y) - Flow_{gt}(x,y)\right\|_1 \cdot \quad (6)$$

In Equation (6), $Flow_{out}(x,y)$ and $Flow_{gt}(x,y)$ represent the estimated and true values of the optical flow at pixel location $(x,y)^T$. $\|\ \|_1$ represents the L1 parametric regularization operation. N is the number of valid pixel points.

To obtain continuous and smooth optical flow field, so as to reduce motion edge noise and discontinuity, this paper adds a smooth term smooth L1 loss function, as in Equation (7).

$$L_{Smooth}(L_{flow\_epe}) = \begin{cases} 0.5x^2, if\ \left|L_{flow\_epe}\right| < 1 \\ \left|L_{flow\_epe}\right| - 0.5, otherwise \end{cases} \cdot \quad (7)$$

Inspired by reference [18], to improve the motion edge blurring and transition smoothing phenomenon of Equation (7), this paper introduces a gradient term based on the endpoint error loss function, so as to better capture the details of the optical flow field, as shown in Equation (8):

$$L_{flow\_\nabla epe} = \frac{1}{N}\sum_{x=0}^{W}\sum_{y=0}^{H}\begin{bmatrix}\left\|\nabla_x\left[Flow_{out}(x,y) - Flow_{gt}(x,y)\right]\right\|_1 \\ +\left\|\nabla_y\left[Flow_{out}(x,y) - Flow_{gt}(x,y)\right]\right\|_1\end{bmatrix}. \quad (8)$$

In Equation (8), $\bigtriangledown$x and $\bigtriangledown$y represent the gradient values of the feature points in the horizontal and vertical directions, respectively. Emphasize the optical flow differences near the image and motion boundaries.

In addition, this study used the binary cross-entropy loss as a consistent occlusion loss function:

$$L_{Occ} = \frac{1}{N}\sum_{x=0}^{W}\sum_{y=0}^{H}\left|Occ_{gt} \bullet \log Occ_{out} + (1 - Occ_{gt}) \bullet \log(1 - Occ_{out})\right|, \quad (9)$$

where Occ_{out} is the occlusion estimated optical flow value and Occ_{gt} is the occlusion true optical flow value.

$$L_{flow} = \lambda_0 L_{flow\_epe} + \lambda_1 L_{Smooth(L_{flow\_epe})} + \\ + \lambda_2 L_{flow\_\nabla epe} + \lambda_3 L_{Occ} \cdot \quad (10)$$

In Equation (10), λ_0, λ_1 λ_2 λ_3 the weights of the different loss functions and the dynamic weighting

method are used to calculate the weight values. The calculation is presented in the pseudo-code of algorithm *L* in Table 1.

Table 1
Algorithm L

Algorithm *L*. Calculate loss function

Input: $Flow_{out}(x,y)$ is an optical flow estimate. $Flow_{gt}(x,y)$ is the real value of optical flow. $Occ_{out}(x,y)$ is to estimate the optical flow value of occlusion. $Occ_{gt}(x,y)$ is occlusion real optical flow value.

Output: The calculated value of the loss function

1. function $Loss\left(Flow_{out}(x,y), Flow_{gt}(x,y), Occ_{out}(x,y), Occ_{gt}(x,y)\right)$
2. $flow\_epe \leftarrow L_{flow\_epe}(Flow_{out}(x,y), Flow_{gt}(x,y))$
3. $flow\_L_{Smooth} \leftarrow L_{Smooth}(L_{flow\_epe})$
4. $flow\_\nabla epe \leftarrow L_{flow\_\nabla epe}(Flow_{out}(x,y), Flow_{gt}(x,y))$
5. $flow\_L_{Occ} \leftarrow L_{Occ}(Occ_{gt}, Occ_{out})$
6. $if\ flow\_epe > flow\_L_{Smooth}$
7. $then\ \lambda_0 \leftarrow 1, \lambda_1 \leftarrow flow\_epe / flow\_L_{Smooth}$
8. $else\ \lambda_0 \leftarrow flow\_L_{Smooth} / flow\_epe, \lambda_1 \leftarrow 1$
9. $if\ flow\_\nabla epe > flow\_epe$
10. $then\ \lambda_2 \leftarrow 1$
11. $else\ \lambda_2 \leftarrow flow\_\nabla epe / flow\_epe$
12. $if\ flow\_L_{Occ} > flow\_epe$
13. $then\ \lambda_3 \leftarrow 1$
14. $else\ \lambda_3 \leftarrow flow\_L_{Occ} / flow\_epe$
15. $return\ \lambda_0 \bullet flow\_epe + \lambda_1 \bullet flow\_L_{Smooth} + \lambda_2 \bullet flow\_\nabla epe + \lambda_3 \bullet flow\_L_{Occ}$

4. Experiment and Data Analysis

In this paper, the MPI Sintel Clean/Final [18], Flying Chairs [5] and KITTI [5] datasets are selected to carry out experiments to evaluate the algorithms proposed in this paper, and the specific parameters of the datasets are given in Table 2. In addition, the value of the bidirectional consistency detection algorithm and the dynamic loss function used for occlusion detection are analyzed through ablation experiments and the algorithms are evaluated for improvement of the endpoint error values.

Table 2
Indicators of specific parameters of the dataset

Datasets	Sizes	Number of training sets	Number of testing sets
MPI Sintel Clean	1920×1080 pixel	1040 pairs	564 pairs
MPI Sintel Final	1024×436 pixel		
Flying Chairs	227×227 pixel	22872 pairs	1175 pairs
KITTI	1226×370 pixel	194 pairs	195 pairs

As shown in Table 3, this study used an NVIDIA Quadro P5000 GPU to test the network on a PyTorch platform with a batch size of 8. Adam [21] is used as the optimization method, and the parameters ß1=0.9, ß2=0.999. The learning rate is set to 1e-5. Train 20k epochs to stop training, and save the training model as a verification model for testing. The total number of network parameters is approximately 149 MB.

Table 3
Training Parameter Settings

Training parameters		Short-cut process
Initial learning rate		0.00001
Batch size		8
Adam	ß1	0.9
	ß2	0.009
Epochs		20k
Total number of parameters		149MB

4.1. Ablation Study

In this section, to prove the effectiveness of the proposed bidirectional consistency and loss function, ablation experiments within a single module and multiple modules are conducted.

4.1.1. The Effectiveness of Bidirectional Consistency Detection

The bidirectional consistency detection modules for occlusion detection consist of two main parts. In the first part, the encoder-decoder takes the consistency of the output bidirectional optical flow value as the

constraint condition of occlusion detection, namely *Occ0*. The second part is based on SSIM occlusion detection by comparing the bidirectional luminance, contrast, and structure consistency as another constraint for occlusion detection, namely *Occ1*.

As shown in Table 4, the EPE values obtained when Occ0 and Occ1 are separately used for occlusion detection in the Sintel Clean data sets are similar, and the EPE values are 3.78 and 3.83 respectively. The EPE value of the combination of the two increased by 22% compared to the EPE value of the no-occlusion module. Figure 5 shows the schematic diagram of the optical flow estimation with different occlusion-detection constraints. As the occlusion-detection constraints are enriched, the arm and background demarcation lines in the red matrix box become clearer.

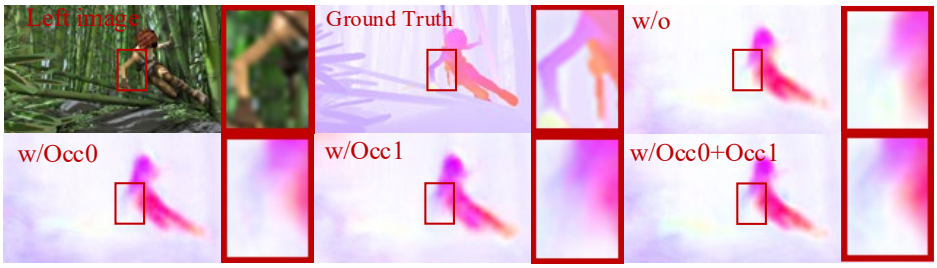
Table 4

Comparison of EPE values with and without bidirectional consistency detection testing. The best EPE values are in bold. Where 'o' is the no-change network, 'Occ0' is the masking result from the encoder-decoder output, and 'Occ1' is the masking information from the SSIM output

Dataset	w/o	w/Occ0	w/Occ1	w/Occ0+Occ1
Sintel Clean	4.50	3.78	3.83	**3.51**
Sintel Final	5.34	4.54	4.45	**4.23**

Figure 5

An example of optical flow estimation with bidirectional consistency detection. The red rectangular box shows the enlarged image of details



4.1.2. Comparison of Different Loss Function Combinations

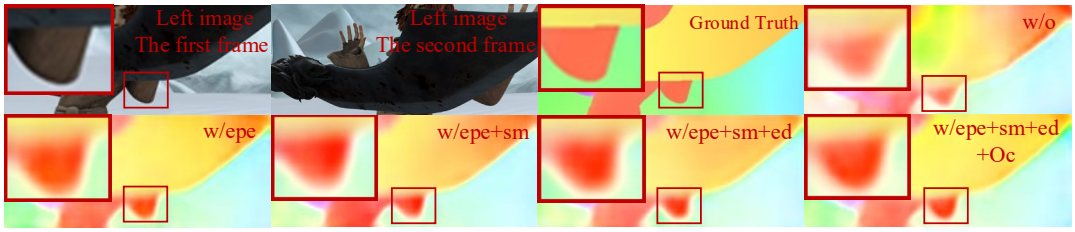
Table 5 shows the superimposed combinations of four loss functions, with the 'w/o' loss function the of L2 paradigm. In the Sintel Clean dataset, the dynamic weighted multi-loss function proposed in this study improved the EPE value by 24.22% compared to the unaltered network. Figure 6 shows the schematic diagram of the optical flow, the overall motion trend of the red matrix is consistent, but the supervised models trained with different loss functions produce different details of the elbow joint edges. The second row and fourth column of this method have smooth and clear edges between the elbow joint edges and the background.

Table 5

Comparison of loss functions for different terms. The best value of EPE is in bold. Where 'o' is the no-change network, 'epe' is the end-point error loss function, 'sm' is the smooth L1 loss, 'ed ' is the gradient loss function, 'Oc' is the occlusion loss function

Dataset	w/o	w/epe	w/epe+sm	w/epe+sm+ed	w/epe+sm+ed+Oc
Sintel Clean	4.50	3.78	3.64	3.61	3.55
Sintel Final	5.34	5.33	5.29	5.25	5.23

Figure 6
Example of multiple loss function optical flow estimation. The red rectangular box is the image after detail enlargement



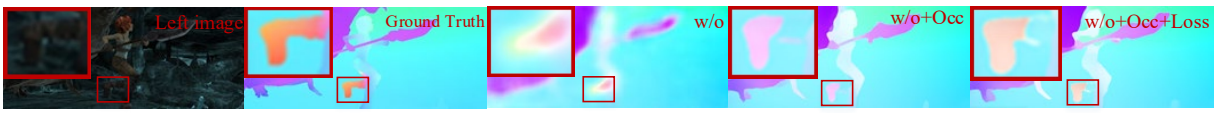
4.1.3. Effectiveness of Different Module Selections

This study proposes two modules: a bidirectional consistency check module for occlusion detection (to for improving object tracking accuracy in occlusion scenes), and a dynamic weight loss function model (for supervised learning of optical flow estimation and occlusion estimation to mitigate the effect of occlusion). As shown in Table 6 and Figure 7, the variation in EPE values and example plots of optical flow estimation with certain modules enabled are illustrated. As modules stack gradually, the training time increases, while the EPE value decreases. The EPE value of the whole network decreased by 3.49 compared with the baseline. As the number of modules increased, the overall trend of the foot movement direction and speed became closer to the real optical flow value.

Table 6
Comparison of multi-module ablation experiments. The black body is the best value of EPE. where 'o' is the change-free network, 'Occ' is the bidirectional consistency checking module for occlusion detection, and 'Loss' is the dynamic weight loss function

Dataset	w/o	w/o+Occ	w/o+Occ+Loss
Sintel Clean	4.50	1.26	**1.01**
Sintel Final	5.34	3.00	**1.07**

Figure 7
Example of multi-module optical flow estimation. The red rectangular box is the image after detail enlargement



4.2. Comparison to the State of the Art

In this section, the proposed FB-Occ method is run on the dataset of MPI-Sintel and Flying Chairs, and the EPE values are compared with FlowNet class algorithm(FlowNetS(2017), FlowNet2(2018), FlowNet3(2019)), supervised algo-rithm(MaskFlow(2020), ScopeFlow(2020), LiteFlowNet2(2020), DICL(2020), RAFT(2020), SCV(2021), PMC-PWC(2021)), and unsupervised algorithm(DDFlow(2019), Epicflow(2019), SelFlow(2019), SimFlow(2020), UFlow(2020), Oc-cInpFlow(2020), UPFlow(2021), FPCR-Net(2022)).

As shown in Table 7, on the Sintel Clean and Final training sets, the proposed method in this paper obtains the best EPE values compared to the optical flow estimation methods in the table, with EPE values of 1.01 and 1.07, respectively. On the Sintel Clean training set, compared to FPCR-Net (unsupervised best value), SCV (supervised best value), and FlowNet3 (best value of the FlowNet family of algorithms), the

EPE values improve by 1.23, 0.28, and 0.46, respectively. On the Sintel Final training set, compared to UPFlow (unsupervised best), PMC-PWC (supervised best), and FlowNet3 (best of the FlowNet family of algorithms), the EPE values improve by 1.60, 1.34, and 1.05, respectively. The EPE value on the Flying Chairs dataset is 0.88. On the KITTI 2012 dataset, the EPE value of the proposed method is superior to most methods, but inferior to the unsupervised SelFlow and OccInpFlow methods. Among them, the EPE value of the proposed method increases by 89.33% compared to the best unsupervised OccInpFlow method and decreases by 17.60% compared to the best supervised FlowNet2 method.

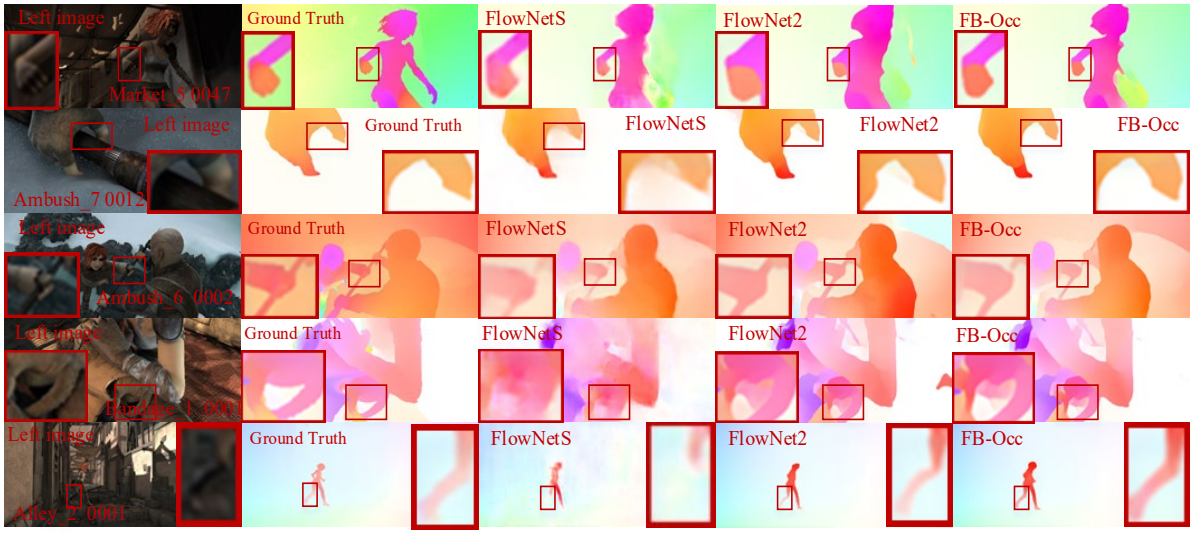
The algorithms used in this study are compared to FlowNetS, as shown in Figure 8. It can be seen that the FB-Occ network proposed in this study achieves better performance and better preservation of the edge detail between the motion-obscured foreground and background. However, the finer edges produced by the movement are not effectively pre-served, and further research will be conducted in the future.

Table 7

Validates the network EPE results on the MPI Sintel, Flying Chairs and KITTI datasets; "-" indicates that results are not reported. The best values for different classifications are in bold

Type	Method	Sintel Clean		Sintel Final		Flying Chairs	KITTI
		EPE train	EPE test	EPE train	EPE test	EPE test	EPE train
Unsupervised	FPCR-Net [37]	2.24	-	3.50	-	-	4.32
	Epicflow [42]	3.54	7.00	4.99	8.51	2.94	2.51
	SelFlow [25]	2.96	6.56	4.06	6.57	.	1.97
	OccInpFlow [28]	2.82	5.79	4.13	7.28	2.62	1.78
	SimFlow [16]	2.86	5.92	3.57	6.92	-	-
	UFlow [19]	2.50	5.21	3.39	6.50	-	-
	UPFlow [27]	2.33	4.68	2.67	5.32	-	-
	DDFlow [24]	2.92	6.18	3.98	7.40	2.97	-
Supervised	FlowNetS [9]	4.50	7.42	5.45	8.43	2.71	8.26
	FlowNet2 [14]	2.02	3.96	3.14	6.02	1.68	4.09
	FlowNet3 [15]	1.47	4.34	2.12	5.67	-	-
	ScopeFlow [3]	3.59	-	4.10	-	-	-
	LiteFlowNet2 [11]	2.24	-	3.78	-	2.63	-
	DICL [36]	1.94	-	3.77	-	-	-
	RAFT [34]	1.43	-	2.71	-	-	-
	SCV [17]	1.29	-	2.95	-	-	-
	PMC-PWC [8]	1.53	3.17	2.41	4.56	-	-
	FB-Occ (Ours)	1.01	3.32	1.07	3.66	0.88	3.37

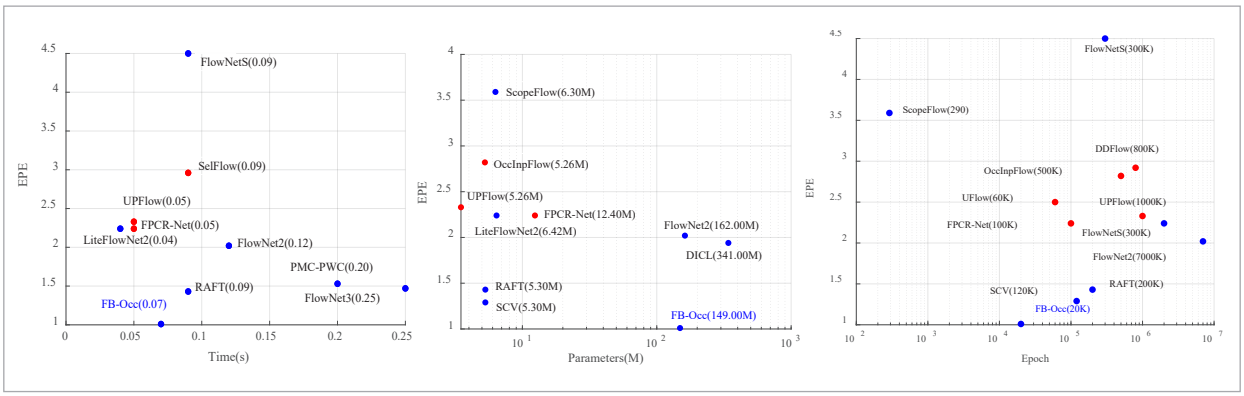
Figure 8
Example of MPI-Sintel optical flow estimation



4.3. Timing and Parameter Counts

As shown in Figure 9, this method is timed on the MPI dataset using an NVIDIA Quadro P5000 GPU. The y-axis represents the EPE value, the red solid circle represents the unsupervised optical flow estimation method, the blue solid circle represents the supervised optical flow estimation method, and FB-Occ is the method proposed in this paper. The time and parameters used to compare FB-OCC are obtained from the original paper on the comparison method, which reported the best parameters. LiteFlowNet2, FPCF-Net, and UPFlow runtimes are superior to those of the algorithm proposed in this study but inferior to those of the algorithm proposed in this paper in terms of training epochs. In terms of parameter calculation, the proposed algorithm is superior to FlowNet2 and DICL. The EPE accuracy is superior to all algorithms. The algorithm in this paper can obtain motion occlusion information in a larger and more accurate range

Figure 9
Shows the comparison of running time, the number of parameters, and several epochs. The y-axis is the EPE value, the red solid circle is the unsupervised optical flow estimation method, the blue solid circle is the supervised optical flow estimation method, and FB-Occ is the method proposed. Example of MPI-Sintel optical flow estimation



while maintaining the occlusion detection accuracy and significantly improving the accuracy of optical flow estimation without significantly in-creasing the time consumption, with the best overall performance.

5. Conclusions

This study proposes an optical flow estimation method based on bidirectional consistency combined occlusion inference to improve the edge blurring problem of motion objects caused by occlusion. First, by swapping two adjacent RGB frames as two inputs, they are simultaneously inputted to an encoder-decoder. Then, the output bidirectional optical flow values are used as constraints on the optical flow and SSIM bidirectional consistency of the occlusion detection, which is used to mitigate the accuracy degradation caused by occlusion. Finally, the supervised training of the optical flow estimation network based on bidirectional consistency joint occlusion inference is completed by using dynamic weight multiple loss function combination. Experiments are carried out on the Sintel Clean/Final training set, Flying Chairs and KITTI dataset. Experimental results show that compared with the state of the art, the proposed method achieves the highest EPE value and the inference time is 14 frames/s. Future research will use lightweight network architectures to enhance real-time processing ability, utilize the advantages of bidirectional consistency and occlusion inference to improve the accuracy of 3D optical flow estimation, and extend this method to 3D scenes, especially for applications in medical imaging or virtual reality. In this way, it helps to solve the motion blurring and occlusion problems in the volume data, improving the effect of 3D reconstruction and accurate diagnosis.

Acknowledgement

Thanks to the Science and technology development plan of Jilin Province for help identifying collaborators for this work.

References

1. Anurag, R. J., Jampani, V., Balles, L., Kim, K., Sun, D., Wulff, J., Black, M. J. Competitive Collaboration: Joint Unsupervised Learning of Depth, Camera Motion, Optical Flow and Motion Segmentation. CVPR, 2019, 3, 12240-12249. https://doi.org/10.1109/CVPR.2019.01252
2. Bailer, C., Varanasi, K., Stricker, D. CNN-Based Patch Matching for Optical Flow with Thresholded Hinge Embedding Loss. CVPR, 2017. https://doi.org/10.1109/CVPR.2017.290
3. Bar-Haim, A., Wolf, L. ScopeFlow: Dynamic Scene Scoping for Optical Flow. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, 7998-8007. https://doi.org/10.1109/CVPR42600.2020.00802
4. Bonneel, N., Tompkin, J., Sunkavalli, K., Sun, D., Paris, S., Pfister, H. Blind Video Temporal Consistency. ACM Transactions on Graphics, 2015, 34(6), 196:1-196:9. https://doi.org/10.1145/2816795.2818107
5. Butler, D., Wulff, J., Stanley, G., Black, M. A Naturalistic Open Source Movie for Optical Flow Evaluation. Proceedings of the European Conference on Computer Vision, 2012, 611-625. https://doi.org/10.1007/978-3-642-33783-3_44
6. Chen, C. F. R., Fan, Q., Panda, R. CrossViT: Cross-Attention Multi-Scale Vision Transformer for Image Classification. IEEE/CVF International Conference on Computer Vision, 2021, 357-366. https://doi.org/10.1109/ICCV48922.2021.00041
7. Chen, J., Cai, Z., Lai, J., Xie, X. Efficient Segmentation-Based PatchMatch for Large Displacement Optical Flow Estimation. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(12), 3595-3607. https://doi.org/10.1109/CRV.2017.40
8. Cza, B., Cheng, F. A., Zhen, C. A., et al. Parallel Multi-scale Context-Based Edge-Preserving Optical Flow Estimation with Occlusion Detection. Image and Vision Computing, 2021. https://doi.org/10.1016/j.image.2021.116560
9. Dosovitskiy, A., Fischer, P., Ilg, E., Häusser, P., Hazirbas, C., Golkov, V., van der Smagt, P., Cremers, D., Brox, T. FlowNet: Learning Optical Flow with Convolutional Networks. International Conference on Computer Vision and Pattern Recognition, 2015, 2758-2766. https://doi.org/10.1109/ICCV.2015.316
10. Ercolino, S., Devoto, A., Monorchio, L., Santini, M., Mazzaro, S., Scardapane, S. On the Robustness of Vision Transformers for In-Flight Monocular Depth Es-

timation. Industrial Artificial Intelligence, 2023, 1(1). https://doi.org/10.1007/s44244-023-00005-3

11. Hui, T.-W., Tang, X., Loy, C. C. A Lightweight Optical Flow CNN-Revisiting Data Fidelity and Regularization. arXiv preprint arXiv:1903.07414, 2019. https://doi.org/10.1109/TPAMI.2020.2976928
12. Hui, T.-W., Tang, X., Loy, C. C. LiteFlowNet: A Lightweight Convolutional Neural Network for Optical Flow Estimation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, 8981-8989. https://doi.org/10.1109/CVPR.2018.00936
13. Hur, J., Roth, S. MirrorFlow: Exploiting Symmetries in Joint Optical Flow and Occlusion Estimation. ICCV, 2017, 1, 312-321. https://doi.org/10.1109/ICCV.2017.42
14. Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., Brox, T. FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks. CVPR, 2017, 1647-1655. https://doi.org/10.1109/CVPR.2017.179
15. Ilg, E., Saikia, T., Keuper, M., Brox, T. Occlusions, Motion and Depth Boundaries with a Generic Network for Disparity, Optical Flow or Scene Flow Estimation. European Conference on Computer Vision, 2018, 614-630. https://doi.org/10.1007/978-3-030-01258-8_38
16. Im, W., Kim, T.-K., Yoon, S.-E. Unsupervised Learning of Optical Flow with Deep Feature Similarity. Proceedings of the European Conference on Computer Vision, 2020. https://doi.org/10.1007/978-3-030-58586-0_11
17. Jiang, S., Lu, Y., Li, H., Hartley, R. Learning Optical Flow from a Few Matches. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, 16592-16600. https://doi.org/10.1109/CVPR46437.2021.01632
18. Jka, B., Lin, C. B., Fei, D. A. Context Pyramidal Network for Stereo Matching Regularized by Disparity Gradients. ISPRS Journal of Photogrammetry and Remote Sensing, 2019, 157, 201-215. https://doi.org/10.1016/j.isprsjprs.2019.09.012
19. Jonschkowski, R., Stone, A., Barron, J. T., Gordon, A., Konolige, K., Angelova, A. What Matters in Unsupervised Optical Flow. arXiv preprint arXiv:2006.04902, 2020. https://doi.org/10.1007/978-3-030-58536-5_33
20. Kennedy, R., Taylor, C. J. Optical Flow with Geometric Occlusion Estimation and Fusion of Multiple Frames. International Workshop on Energy Minimization Methods in Computer Vision & Pattern Recognition, Springer International Publishing, 2015. https://doi.org/10.1007/978-3-319-14612-6_27
21. Kingma, D. P., Ba, J. Adam: A Method for Stochastic Optimization. ICLR, 2015. https://doi.org/10.1201/9781003240167-10
22. Kroeger, T., Timofte, R., Dai, D., Van Gool, L. Fast Optical Flow Using Dense Inverse Search. In Proceedings of ECCV, 2016, 471-488. https://doi.org/10.1007/978-3-319-46493-0_2
23. Lai, W.-S., Huang, J.-B., Yang, M.-H. Semi-Supervised Learning for Optical Flow with Generative Adversarial Networks. In Proceedings of NeurIPS, 2017, 354-364. https://doi.org/10.31274/etd-180810-6069
24. Liu, P., King, I., Lyu, M., Xu, J. DDFlow: Learning Optical Flow with Unlabeled Data Distillation. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 8770-8777. https://doi.org/10.1609/aaai.v33i01.33018770
25. Liu, P., Michael, L., King, I., Xu, J. SelFlow: Self-Supervised Learning of Optical Flow. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, 4571-4580. https://doi.org/10.15607/RSS.2005.I.036
26. Lu, Y., Valmadre, J., Wang, H. DEVON: Deformable Volume Network for Learning Optical Flow. IEEE/CVF Winter Conference on Applications of Computer Vision, 2020, 2705-2713. https://doi.org/10.1109/WACV45572.2020.9093590
27. Luo, K., Wang, C., Liu, S. UPFlow: Upsampling Pyramid for Unsupervised Optical Flow Learning. CVPR, 2021. https://doi.org/10.1109/CVPR46437.2021.00110
28. Luo, K., Wang, C., Ye, N. OccInpFlow: Occlusion-Inpainting Optical Flow Estimation by Unsupervised Learning. arXiv preprint arXiv:2006.16637, 2020. DOI: 10.48550/arXiv.2006.16637.
29. Menze, M., Geiger, A. Object Scene Flow for Autonomous Vehicles. CVPR, 2015, 1(7), 3061-3070. https://doi.org/10.1109/CVPR.2015.7298925
30. Ranjan, A., Black, M. J. Optical Flow Estimation Using a Spatial Pyramid Network. CVPR, 2017, 2720-2729. https://doi.org/10.1109/CVPR.2017.291
31. Sevilla-Lara, L., Sun, D. Q., Jampani, V., Black, M. J. Optical Flow with Semantic Segmentation and Localized Layers. CVPR, 2016, 1, 3889-3898. https://doi.org/10.1109/CVPR.2016.422
32. Sun, D., Liu, C., Pfister, H. Local Layering for Joint Motion Estimation and Occlusion Detection. CVPR, 2014. https://doi.org/10.1109/CVPR.2014.144
33. Sun, D., Yang, X., Liu, M.-Y., Kautz, J. PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume. CVPR, 2018, 8934-8943. https://doi.org/10.1109/CVPR.2018.00931
34. Teed, Z., Deng, J. RAFT: Recurrent All-Pairs Field Transforms for Optical Flow. ECCV, 2020. https://doi.org/10.24963/ijcai.2021/662

35. Tummala, S., Kadry, S., Bukhari, S. A. C., et al. Classification of Brain Tumor from Magnetic Resonance Imaging Using Vision Transformers Ensembling. Current Oncology, 2022, 29(10), 7498-7511. https://doi.org/10.3390/curroncol29100590

36. Wang, J., Zhong, Y., Dai, Y., Zhang, K., Ji, P., Li, H. Displacement-Invariant Matching Cost Learning for Accurate Optical Flow Estimation. NeurIPS, 2020. https://doi.org/10.1109/ICCV.2015.457

37. Xiaolin, S., Yuyang, Z., Jingyu, Y., Cuiling, L., Wenjun, Z. FPCR-Net: Feature Pyramidal Correlation and Residual Reconstruction for Optical Flow Estimation. Neurocomputing, 2022, 417, 346-357. https://doi.org/10.1016/j.neucom.2021.11.037

38. Yang, F., Su, L., Zhao, J., Chen, X., Wang, X., Jiang, N., Hu, Q. SA-FlowNet: Event-Based Self-Attention Optical Flow Estimation with Spiking-Analogue Neural Networks. IET Computer Vision, 2023, 1-11. https://doi.org/10.1049/cvi2.12206

39. Zhai, M., Ni, K., Xie, J., Gao, H. Scene Flow Estimation from 3D Point Clouds Based on Dual-Branch Implicit Neural Representations. IET Computer Vision, 2023, 1-14. https://doi.org/10.1049/cvi2.12237

40. Zhong, Y., Ji, P., Wang, J., Dai, Y., Li, H. Unsupervised Deep Epipolar Flow for Stationary or Dynamic Scenes. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, 12095-12104. https://doi.org/10.1109/CVPR.2019.01237

41. Zhao, B., Huang, Y., Wei, H., Hu, X. Ego-Motion Estimation Using Recurrent Convolutional Neural Networks Through Optical Flow Learning. Electronics, 2021, 10(3), 222. https://doi.org/10.3390/electronics10030222

42. Zhao, S., Sheng, Y., Dong, Y., Chang, E. I., Xu, Y., et al. MaskFlowNet: Asymmetric Feature Matching with Learnable Occlusion Mask. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, 6278-6287. https://doi.org/10.1109/CVPR42600.2020.00631

43. Zheng, Y., Jiang, W. Evaluation of Vision Transformers for Traffic Sign Classification. Wireless Communications and Mobile Computing, 2022, 1, 3041117. https://doi.org/10.1155/2022/3041117