Computer Vision and Image Understanding 249 (2024) 104160

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

# Estimating optical flow: A comprehensive review of the state of the art

Andrea Alfarano [a], Luca Maiano [a,b,*], Lorenzo Papa [a], Irene Amerini [a]

[a] Sapienza University of Rome, Via Ariosto, 25, 00185, Roma RM, Italy
[b] Ubiquitous srl, Via Ariosto, 25, 00185, Roma RM, Italy

ARTICLE INFO

Communicated by R. Venkatesh Babu

*MSC:*
41A05
41A10
65D05
65D17

*Keywords:*
Computer Vision
Optical flow
motion estimation

ABSTRACT

Optical flow estimation is a crucial task in computer vision that provides low-level motion information. Despite recent advances, real-world applications still present significant challenges. This survey provides an overview of optical flow techniques and their application. For a comprehensive review, this survey covers both classical frameworks and the latest AI-based techniques. In doing so, we highlight the limitations of current benchmarks and metrics, underscoring the need for more representative datasets and comprehensive evaluation methods. The survey also highlights the importance of integrating industry knowledge and adopting training practices optimized for deep learning-based models. By addressing these issues, future research can aid the development of robust and efficient optical flow methods that can effectively address real-world scenarios.

## 1. Introduction

Optical flow enables motion understanding by examining pixel displacements across a sequence of frames. This method estimates the 2D projections of 3D point motion on the camera plane, determining pixel shifts between consecutive images in a sequence, and holds significant importance in computer vision thanks to its applicability to various higher-level tasks that range from separating moving objects from backgrounds (Brox et al., 2004a) to object identification and tracking (Bailer et al., 2017) as well as approximating the three-dimensional structure of a scene (Vedula et al., 1999).

The seminal formulation of optical flow has been proposed by Gibson (1950), but it took nearly three decades for the development of the first frameworks capable of computing dense optical flow fields. Horn and Schunck (1981) proposed the first pioneering variational framework that minimizes an energy function composed of a data term and a regularization term to ensure smoothness in the estimated motion field. Over time, numerous approaches have been devised to improve the accuracy and robustness of the estimated flow field by incorporating handcrafted features and strategies. Section 3 will review all these advances in detail.

While classical optical flow computation frameworks have made substantial advancements, their application has often been limited due to the significant computational demands and constrained performance (Shah et al., 2021). On the contrary, the advent of deep learning marked the beginning of significant progress. Deep learning models can directly estimate optical flow from image pairs, overcoming the need to use handcrafted features that are often too expensive to extract and poorly generalizable, thus introducing important improvements in both accuracy and computational efficiency (Ilg et al., 2017). In this respect, FlowNet (Fischer et al., 2015) represented a pioneering optical flow deep learning model. Following this approach, Ilg et al. (2017) enhanced the FlowNet by stacking multiple modules, leading to the creation of FlowNet 2.0. This marked a significant milestone as it was the first instance where a deep learning model exceeded the performance of classical algorithmic methods.

Interestingly, the integration of classical methods with deep learning has proved to be a promising direction. In fact, classical methods still play a fundamental role in modern advancements and continue to stimulate innovation in this field (Shah et al., 2021; Savian et al., 2020). For instance, Ranjan and Black (2017) combined the traditional spatial-pyramid framework with deep learning, resulting in a model that matched the effectiveness of FlowNet (Fischer et al., 2015) but with a remarkable 96% reduction in the number of parameters. From here, Sun et al. (2018) further incorporated classical techniques such as pyramidal processing, warping, and cost volume calculations to create PWC-Net. This model outperformed FlowNet 2.0 with a significantly smaller architecture, approximately 17 times lighter. Similarly, Hui et al. (2018) have introduced a network (LiteFlowNet) which, by

* Corresponding author.
*E-mail addresses:* alfarano.2026770@studenti.uniroma1.it (A. Alfarano), maiano@diag.uniroma1.it (L. Maiano), papa@diag.uniroma1.it (L. Papa), amerini@diag.uniroma1.it (I. Amerini).

https://doi.org/10.1016/j.cviu.2024.104160
Received 27 November 2023; Received in revised form 27 May 2024; Accepted 4 September 2024
Available online 16 September 2024

combining pyramid processing with flow regularization, maintains a comparable accuracy despite being 30 times smaller than FlowNet 2.0 and 1.36 times faster. Moreover, in addition to augmenting model performance, embedding optical flow knowledge into neural network algorithms opens up the way for unsupervised and semi-supervised training methods, thereby reducing the dependence on labeled data (Tu et al., 2019; Dobrički et al., 2022b; Yu et al., 2016)

The enormous development of these methodologies has led to various surveys, each focusing on different aspects of the field. Fortun et al. (2015) present a survey on classical optical flow computational frameworks, (Shah et al., 2021), (Tu et al., 2019), and (Savian et al., 2020) discuss both classical and neural network-based methods. Dobrički et al. (2022b) concentrate on unsupervised learning methods, Zhai et al. (2021) explore optical flow alongside the broader concepts of 2.5D and 3D scene flow. Guney et al. (2018) outline various fields of application, and Mathur (2020) provide an extensive survey of optical flow datasets. Previous surveys have thoroughly examined the theoretical foundations of optical flow methods, but a crucial gap remains in understanding the practical challenges encountered in real-world scenarios. While modern deep learning-based approaches often excel on standardized benchmarks, their effectiveness in diverse real-world settings remains uncertain. To bridge this gap, this survey systematically investigates the most representative challenges involved in optical flow computation, providing insights into developing more robust and adaptable models. Recognizing the demand for real-time computation in many applications, we examine classical and current strategies to accelerate processing while maintaining performance. In this survey, we also take into account the added challenges presented by training and assessing deep learning models. Acquiring real-world data with accurate optical flow ground truth is a current challenge, often leading researchers to depend on generic synthetic datasets. However, these datasets often struggle to effectively transfer knowledge to real-world, task-specific situations. To address this limitation, our literature review provides the largest collection of large optical flow datasets to date, helping researchers identify the most suitable datasets for their needs. Additionally, the black-box nature of deep learning models can raises concerns about their evaluation, particularly in relation to specific challenges and project needs. This survey directly addresses these concerns, providing insights into evaluating performance in the context of real-world applications.

The rest of this work is organized as follows: in Section 2, we first go over some important optical flow methods and identify the most important applications of optical flow, presenting a large collection of examples. In Section 3, we explore the main problems faced in optical flow computation, such as occlusion, large displacements, changes in light, and computational constraints, while also presenting the main methods used to deal with these problems, focusing on the assumptions and strategies used in classical approaches. In Section 4, we go into the process of evaluating methods of computation, providing a detailed analysis of both traditional and advanced ways to measure performance. In Section 5, we present the most comprehensive collection of optical flow datasets by the knowledge of the authors. Lastly, in Section 6, we look at potential future research directions in optical flow computation and bring up some unanswered questions in the optical flow literature and its application to real-world situations, and draw our conclusion in Section 7.

## 2. Preliminaries

In this section, we review the evolution of the leading optical flow techniques from classical algorithmic methods to transformer-based deep learning models and discuss the most relevant fields in which these techniques have been applied. To guarantee a complete overview of the main aspects raised by previous studies, in Section 2.1, we will retrace the evolution of optical flow methods, providing an overview ranging from variational methods to transformer-based models. Subsequently, in Section 2.2, we will see the most relevant fields of application for optical flow.

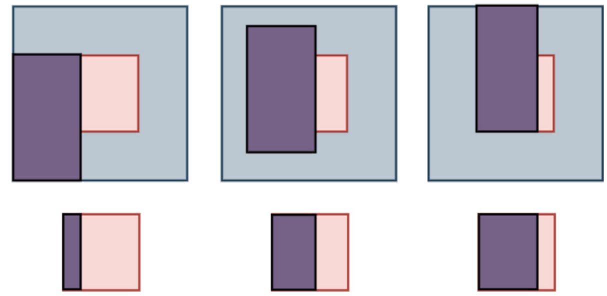

**Fig. 1.** Depiction of the aperture problem. Upper row: Actual movement of the purple rectangle. Lower row: Perceived movement of the purple rectangle on the smaller area (red), where only the horizontal component of the motion is evident. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

### 2.1. Progress in optical flow estimation techniques

The field of optical flow estimation has seen significant progress since its introduction in the 1950s (Gibson, 1950). The assumption that makes estimating these flow vectors possible is the brightness constancy constraint, which states that the intensity of a pixel remains constant as it moves. Formally, this is expressed as:

$$\frac{\partial I}{\partial t} + u\frac{\partial I}{\partial x} + v\frac{\partial I}{\partial y} = 0, \tag{1}$$

where $I$ represents the image intensity, $u$ and $v$ are the horizontal and vertical components of the optical flow vector, and $\frac{\partial I}{\partial t}$, $\frac{\partial I}{\partial x}$, and $\frac{\partial I}{\partial y}$ are the temporal and spatial derivatives of $I$.

However, this constraint alone introduces the so-called *aperture problem*, depicted in Fig. 1, arising from the inherent ambiguity in determining an object's motion based solely on information observed within a small local region (i.e., the "aperture"). Specifically, attempting to solve for both $u$ and $v$ using only Eq. (1) results in an underdetermined system. To resolve this ambiguity and obtain a well-posed problem, additional constraints or assumptions must be introduced.

Specifically, we can introduce two variational frameworks for computing dense optical flow from consecutive frames: the global *HS* method by Horn and Schunck (1981), which uses a global smoothness constraint, and the local *LK* method by Lucas and Kanade (1981), which instead introduce local constancy.

The HS method addresses the aperture problem and minimizes distortions in optical flow by introducing a global constraint of smoothness, which imposes smooth flow variation between neighboring pixels. It formulates the optical flow as a global energy functional to be minimized:

$$E = \iint \left[ (I_x u + I_y v + I_t)^2 + \alpha^2 (\|\nabla u\|^2 + \|\nabla v\|^2) \right] dx\,dy \tag{2}$$

where $I_x$, $I_y$, and $I_t$ are gradients of the image intensity along the x, y, and time dimensions, respectively, $u(x,y)$ and $v(x,y)$ form the optical flow vector at position $(x,y)$, $\alpha$ is a regularization constant that controls the balance between data fidelity and smoothness in the estimated flow, and $\nabla u$ and $\nabla v$ denote the spatial gradients of the optical flow components.

In contrast, the *LK* method assumes that the flow is essentially constant in a local neighborhood and solves the optical flow equations for all pixels within that neighborhood. The brightness constancy assumption leads to the following equation:

$$I_x(p) \cdot V_x + I_y(p) \cdot V_y = -I_t(p). \tag{3}$$

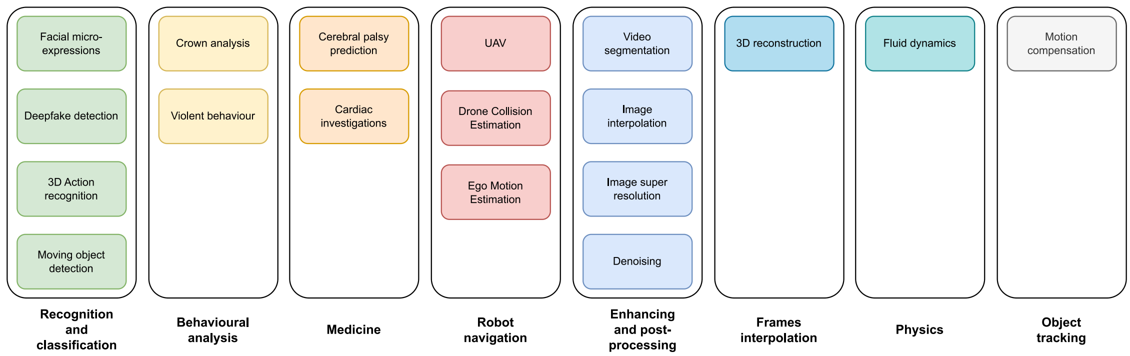


**Fig. 2.** Relevant application fields of Optical flow estimation.

By aggregating data from nearby pixels $(q_1, q_2, \dots, q_n)$, the LK method sets up a system of linear equations $Av = b$:

$$A = \begin{bmatrix} I_x(q_1) & I_y(q_1) \\ I_x(q_2) & I_y(q_2) \\ \vdots & \vdots \\ I_x(q_n) & I_y(q_n) \end{bmatrix}, v = \begin{bmatrix} V_x \\ V_y \end{bmatrix}, b = \begin{bmatrix} -I_t(q_1) \\ -I_t(q_2) \\ \vdots \\ -I_t(q_n) \end{bmatrix}. \tag{4}$$

As this system is generally overdetermined, the LK method uses the least squares principle, solving for $v$ as:

$$v = (A^T A)^{-1} A^T b. \tag{5}$$

Despite these fundamental contributions, the Horn-Schunck and Lucas-Kanade methods introduce several challenges in handling complex scenarios, leading to continued research and advancement in optical flow estimation techniques. For example, the (Brox et al., 2004a) method integrated high-order pyramid smoothness constraints to preserve motion boundaries and improve optical flow estimation in complex scenes. Another example is DeepFlow (Weinzaepfel et al., 2013) and EpicFlow (Revaud et al., 2015), which combine traditional variational frameworks with matching by correspondence to estimate motion displacement between two consecutive frames. They work by first analyzing the features that make up the input frames and then by matching the correspondence (Shah et al., 2021).

The advent of deep learning led to the development of FlowNet (Fischer et al., 2015), which introduced two groundbreaking architectures for optical flow estimation: the FlowNet-S and the FlowNet-C. These architectures are based on the encoder–decoder's U-Net (Ronneberger et al., 2015) structure and facilitate efficient feature extraction and motion estimation. The FlowNet 2.0 (Ilg et al., 2017) builds upon the original FlowNet architecture by stacking multiple FlowNet sub-networks and registered a significant improvement in state of art accuracy, although its application remains challenging in resource-constrained environments (Hui et al., 2018). Differently, the SPyNet (Ranjan and Black, 2017) architecture offered a compact network for optical flow estimation, combining the coarse-to-fine approach of traditional methods with deep learning techniques. Despite its reduced size, its accuracy and ability to handle large motions were limited compared to FlowNet 2.0. The PWC-Net (Sun et al., 2018) proposed a modified spatial pyramid network for optical flow estimation, integrating traditional stereo matching, feature extraction, and cost volume with deep learning techniques. The RAFT (Teed and Deng, 2020) network introduced an innovative method using recurrent all-pairs field transforms and a recurrent update operator, deviating from traditional spatial pyramid networks and offering a unique perspective on optical flow estimation.

With the growing prominence of transformers in computer vision tasks, their application to optical flow computation presents exciting prospects (Vaswani et al., 2017). Transformer-based optical flow models, such as FlowFormer (Huang et al., 2022), utilize self-attention mechanisms to capture long-range dependencies and spatial relationships between pixels, which is essential for accurate optical flow estimation.

### 2.2. Fields of application

This section discusses the most relevant applications of optical flow, as summarized in Fig. 2.

**Recognition and Classification**. Optical flow provides valuable information for recognition and classification tasks. For example, Zeng et al. (2023) conducted a comprehensive survey on the use of optical flow for micro-expression recognition, emphasizing the potential of optical flow in detecting subtle facial movements essential for emotion recognition and human–computer interaction. Similarly, optical flow methods have been applied in Deepfake detection (Jiang et al., 2019; Rössler et al., 2019) by analyzing inconsistencies in the temporal changes of pixel intensities between consecutive generated frames. The generation process can, in fact, introduce subtle discrepancies in the motion of facial features (Li and Lyu, 2018; Sabir et al., 2019; Amerini et al., 2019; Caldelli et al., 2021; Matern et al., 2019; Agarwal et al., 2020). Ren et al. (2018) investigated the application of optical flow in 3D action recognition, demonstrating the utility of optical flow in extracting critical motion information from 3D skeletal data. This integration results in more accurate and robust action recognition across diverse scenarios. Finally, Agarwal et al. (2016) reviewed the role of optical flow in enhancing the robustness and efficiency of moving object detection. They demonstrated that incorporating optical flow into the process could improve the detection and tracking of moving objects in complex and dynamic environments.

**Behavioral Analysis**. Optical flow techniques have found applications in crowd analysis, providing insights into crowd dynamics and behavior (Kajo et al., 2015). Additionally, optical flow has been employed for the detection of abnormal or violent behavior in crowd settings, contributing to improved safety and security measures (Huang and Chen, 2014; Gkountakos et al., 2020).

**Medical Applications**. Optical flow techniques have been extensively applied in the medical domain, contributing to various diagnostic and therapeutic processes. For instance, researchers have used optical flow to predict infantile cerebral palsy by analyzing motion patterns in recorded movements (Stahl et al., 2012). Moreover, the optical flow has facilitated neural investigations (Huang et al., 2016) and the measurement of strain in myocardium tissue (Xu et al., 2010). The assessment of organ movements and dynamics is another area where optical flow has shown significant potential. Researchers have employed optical flow for the examination of the vocal tract (Andrade-Miranda et al., 2018), the evaluation of the respiratory system (Benameur et al., 2017), and the study of other organ dynamics (Angelini and Gerard, 2006). Additionally, optical flow has been instrumental in compensating for organ movements during magnetic resonance therapy (Zachiu et al., 2015).

**Robot Navigation**. Optical flow has emerged as a fundamental technique for mapless robot navigation, playing an important role in estimating the direction and velocity of a robot as it moves from its starting point to a target destination. The inspiration for employing

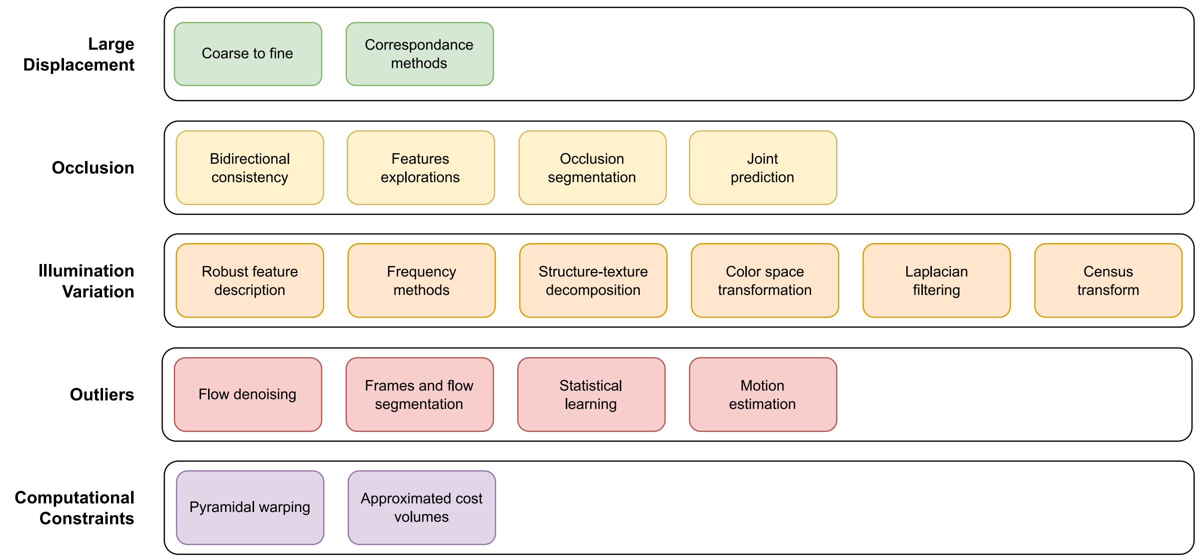


**Fig. 3.** Challenges faced during optical flow estimation and related relevant methods.

optical flow in robot navigation stems from the observation of bees' flying behavior, where they navigate complex environments using visual motion cues. Optical flow has been successfully implemented in various scenarios, such as Unmanned Aerial Vehicles (UAVs) and robotic navigation systems (Chao et al., 2014, 2013). For instance, Editya et al. (2022) demonstrated the utility of optical flow in drone collision estimation, improving the safety and reliability of UAV operations in crowded or dynamic environments. Additionally, Raudies and Neumann (2012) explored the use of optical flow for ego-motion estimation, which is vital for understanding the robot's movement relative to its surroundings and adjusting its trajectory accordingly.

**Enhancing and Post-Processing**. Image interpolation and super-resolution necessitate precise and detailed pixel alignment. Optical flow techniques have been employed in various image-enhancing and post-processing pipelines. Optical flow is used for motion compensation in video super-resolution (Tu et al., 2022a) (Makansi et al., 2017) and video segmentation (Anthwal and Ganotra, 2019; Xiao and Lee, 2016). Some video super-resolution methods used optical flow to improve surveillance recordings (Tu et al., 2022b; Lin et al., 2005). Additionally, optical flow has been utilized for video denoising purposes, resulting in enhanced image quality (Fassold, 2022).

**Frames Interpolation**. Optical flow can be used in general for Image Interpolation of two consecutive images in an image sequence (Chen and Lorenz, 2012). For example, it has been utilized for 3D reconstruction from monocular sequences, enabling the creation of detailed 3D models from 2D image data (Zhang et al., 2016)

**Physical Applications**. Optical flow has emerged as a valuable tool for estimating fluid dynamics, providing non-intrusive measurements that do not alter the flow patterns (Osman et al., 2016; Mendes et al., 2022). This non-contact approach enables the analysis of various fluid systems, even in complex and challenging environments. For instance, Osman and Ovinis (2020) reviewed the application of optical flow methods to estimating oil spill flow rates in intricate deep-water settings. In another application, Radhakrishnan et al. (2017) focused on river flow measurement using video recordings of the water surface, eliminating the need for in-situ measurements and making it particularly suitable during flood events. Finally, Ghalenoei et al. (2014) employed optical flow techniques to estimate sea surface currents.

**Object Tracking**. Optical flow can be used for object tracking in various applications (Yin et al., 2016; Husseini, 2017). By estimating the displacement of the tracked object between consecutive frames, optical flow enables the accurate determination of the object's position and movement over time (Agarwal et al., 2016).

## 3. Challenges and strategies

From the initial formulation of optical flow in 1950 (Gibson, 1950), it took almost three decades to develop the first computational framework (Horn and Schunck, 1981). The difficulties associated with optical flow estimation are mainly due to several challenges that estimation systems must efficiently address to be applied in the real world. Fig. 4 gives an overview of those challenges on MPI-Sintel (Butler et al., 2012) a popular animation benchmark created for replicating such challenges in a controlled environment. For example, changes in a scene's lighting conditions can change the brightness of displaced pixels without implying actual motion, leading to incorrect matches. Furthermore, it is not uncommon in real-world settings for one object to be partially hidden behind another, making it impossible to establish a direct pixel match. Additionally, optical flow methods must demonstrate resilience to noise and other imaging artifacts in real-world data.

All these challenges need to be addressed not only effectively but also efficiently. Optical flow, which provides low-level motion information for high-level vision tasks, is often used in vision pipelines where fast response times, minimal resource usage, and real-time processing are crucial requirements.

Before the advent of end-to-end computation enabled by Neural Networks (NNs), these estimation challenges were managed using meticulously crafted features and strategies grounded in domain knowledge. These approaches necessitated a profound understanding of the problem and its efficient integration into an algorithmic framework. In this section, we examine these significant challenges and suggest a few notable methods, as depicted in Fig. 3. In Section 3.1, we discuss the *large displacement* challenges, i.e., the difficulty in estimating correspondences between objects that are far apart due to rapid motion on the image plane. In Section 3.2, we address the problem of *occlusion*, that is, the complexity of estimating correspondences between obscured or hidden parts of objects. Section 3.3 focuses on the *illumination variation* challenge, which concerns the struggle to estimate optical flow while maintaining robustness to the varying illumination conditions of the scene. Finally, Section 3.4 covers the difficulties in creating methods that can withstand variations in the input frames (like noise) or that can self-detect failures in their predictions, such as unrealistic motion models or smoothed discontinuities in the optical flow maps.

### 3.1. Large displacement

Handling large displacements is a critical issue in real-world optical flow applications, arising from factors such as rapidly moving objects,


**Fig. 4.** Representation of some challenges faced during optical flow estimation on the MPI-Sintel benchmark (Butler et al., 2012), with actual ground truth provided. (A) Large Displacement: Fast-moving objects causing displacements, creating difficulties in finding correspondences. (B) Occlusion: Objects in the scene blocking each other, leading to missing or ambiguous flow information. (C) Illumination Variation: Changes in lighting conditions affecting the consistency of flow estimation. (D) Outliers: Presence of motion blur and fog potentially introducing noise and inaccuracies in flow calculation.

low frame rates, and objects that are close to the camera. Establishing correspondences between distant pixels is a complex task, and many high-performing classical methods have developed strategies to address large displacements. The importance of managing large displacements is also remarked by the characteristics of well-known benchmark datasets such as KITTI (Geiger et al., 2012) and MPI-Sintel (Butler et al., 2012), which often feature vast pixel displacements, sometimes exceeding 100 pixels.

**Coarse-to-Fine Pyramidal Methods**. Coarse-to-fine pyramidal methods, popularized by Brox et al. (2004a), serve as a crucial tool for dealing with large displacement scenarios in optical flow computation. The approach calculates optical flow at multiple resolution scales, starting from a downsampled, lower-resolution version of the original frames. This allows for the initial capture of large motion vectors in a much easier and computationally efficient manner due to the reduced image complexity. This coarse match provides a rough estimate of the flow field and serves as the starting point for iterative refinement. The process then incrementally refines the match, using information from coarser levels to guide the flow estimation at the finer scales. Pyramidal Methods ensure that large displacements, which might be difficult to capture directly at the original resolution, are taken into account from the outset and refined over iterations. However, this method has some drawbacks. As highlighted by Brox et al. (2004a), small objects subjected to large displacements may become indiscernible at coarser scales due to the blurring effects that are introduced by downsampling. Furthermore, Xu et al. (2011) pointed out that even small displacement motion could be affected by this issue. Overlaying layers of motion at coarser levels can lead to inaccuracies and errors, which could propagate and amplify as the computation proceeds at finer scales. Pyramidal methods to handle large displacements are used in notable neural networks like Raft (Teed and Deng, 2020) and PWC-Net (Sun et al., 2018).

**Correspondence Methods**. Optical flow computation can be interpreted as discerning correspondences between points across frames during motion-related shifts. Tu et al. (2019) details two main approaches to this matching problem: (1) *patch-based* and (2) *feature-based*. The patch-based approach, such as PatchMatch (Barnes et al., 2009), finds matches between patches of two image regions. Another architecture following this approach is DeepFlow (Weinzaepfel et al., 2013), which provides robustness to large displacements and appearance changes by leveraging information from surrounding pixels. On the other hand, feature-based approaches like EpicFlow (Revaud et al., 2015) rely on sparse matching algorithms to pinpoint correspondences between key points in two images that are then interpolated to produce a dense optical flow field. Although effective on regions with unique characteristics, these techniques can introduce ambiguities when applied to regions with repetitive textures or lacking distinct features (Bailer et al., 2017; Trinh and Daul, 2019). Furthermore, the exhaustive search between all feature matches can be computationally expensive, presenting a significant challenge for real-time applications or large datasets.

Several strategies have been proposed to enhance both the descriptor and matching stages. These include (1) improving the initial estimated motion field using the nearest neighbor field (Chen et al., 2013), (2) using a regular grid as a heuristic to speed up the correspondence search (Chen and Koltun, 2016), and (3) employing a hierarchical tree approach (Bailer et al., 2019). DeepMatching (Weinzaepfel et al., 2013) uses multiscale scoring of matches to overcome the lack of distinctiveness in small patches and weak textures. Finally, some studies propose hybrid CNN-based matching techniques that leverage CNNs for robust feature descriptors along with traditional matching search algorithms (Bailer et al., 2017).

State-of-the-art deep learning models, including PWC-Net (Sun et al., 2018), Raft (Teed and Deng, 2020), and effectively employ correspondence-based methods using conventional cost volumes to estimate the displacement between frame pixels.

### 3.2. Occlusion

Occlusion is a common occurrence in realistic scenes that can cause errors in optical flow estimation. It arises when a pixel in one image does not have a correspondence in the other image. This can be caused when a moving point gets occluded behind an object due to camera motion or self-occlusion due to an object changing its orientation. Occlusion violates many of the classical optical flow assumptions and is especially challenging when the scene also presents large displacement between moving objects. Some of the most successful classical optical flow methods utilize occlusion as supplementary evidence to compute optical flow (Shah et al., 2021).

**Bidirectional Consistency**. This technique is devised to tackle the problem of verifying the forward and backward flow between a pair of frames, ensuring flow coherence in both directions within the realm of optical flow estimation (Jeong et al., 2022). By comparing the forward flow from the first frame to the second frame with the backward flow from the second frame to the first, inconsistencies can be effectively identified. If the consistency falls below a predefined threshold, the algorithm perceives the region as occluded. In addition, bidirectional consistency can aid in refining flow estimation by reducing the errors and ambiguities associated with occlusions (Bailer et al., 2019). This culminates in more reliable and precise motion estimations. Notably, the state-of-the-art performance achieved by VideoFLows (Shi et al., 2023) is attributed to its utilization of a TRi-frame Optical Flow module (TROF), a technique that jointly estimates bi-directional optical flows for three consecutive frames in videos, further enhancing the understanding of complex motion patterns.

**Features Exploitation**. These methods detect occlusion by using frames and optical flow features (Tu et al., 2019). For instance, EpicFlow by Revaud et al. (2015) uses edge information to refine the flow, and occlusion is inferred from the mismatches between the initial estimate and the refined flow. Differently, the Occlusion-Net (Dinesh Reddy et al., 2019) uses a graph encoder to classify invisible edges and a graph decoder network to correct the occluded keypoint locations.

**Occlusion Detection and Joint Prediction**. Occlusion detection involves three steps: (1) estimating the optical flow while disregarding occlusion, (2) identifying occlusion areas using the (unreliable) estimated optical flow, and then, (3) correcting the optical flow using the computed occlusion areas (Ince and Konrad, 2008). A modern variation of this technique jointly computes occlusion maps and the optical flow fields, sharing useful information during the calculations. Ince and Konrad (2008) propose a variational formulation capable of implicitly detecting occlusion and extrapolating optical flow in occluded areas using anisotropic diffusion. The MaskFlowNet (Zhao et al., 2020) employs an encoder–decoder architecture complemented by an additional mask branch for occlusion prediction. The MirrorFlow network (Janai et al., 2018) applies a symmetry constraint. This method is based on the assumption that a scene is observed from multiple viewpoints between different frames and that the motion between these viewpoints follows a piecewise rigid model. Finally, the OAFlow model (Wang et al., 2018) incorporates a photometric loss term and an occlusion regularizer in its loss function to model occlusion explicitly by leveraging unsupervised learning.

**Energy Minimization and Regularization**. This strategy leverages energy minimization methods that promote spatial continuity and regularization terms that model occlusion (Zach et al., 2007). The main idea is that the occlusion areas will introduce abrupt changes in the optical flow. Thus, energy minimization and regularization methods can be used to detect these changes and smooth out the flow in occluded areas.

### 3.3. Illumination variation

Popular classical optical flow frameworks assume that the intensity or brightness of a pixel remains constant between consecutive frames in a sequence of images or a video (Horn and Schunck, 1981). This classical assumption is called *Brightness Consistence Assumption* (BCA), and in other words, when an object moves within the field of view, the brightness of the object's pixels is expected to stay the same as it moves from one frame to the next. This simplifies the optical flow computational problem by estimating the displacements of the intensity pattern rather than the true motion of the pixels between two pictures. However, the motion of an object between frames is not solely correlated to its brightness path. In real-world scenarios, changes in lighting conditions are frequent due to factors such as the sun, the clouds, the transitions from lights to shadows, or moving a light source in front of a stationary object, which will produce changes in brightness without moving any object in the scene.

**Robust Feature Description**. These approaches mitigate false matches resulting from feature descriptors that are sensitive to illumination changes. Trinh and Daul (2019) propose a local illumination change model that mathematically verifies whether a descriptor is invariant or susceptible to illumination variations between images. The same work also introduces two general mathematical formulations, which can be applied to devise a range of new illumination-invariant descriptors. Mohamed et al. (2014) suggest an illumination-robust optical flow method. In their approach, the data term of the variational model is extracted from texture features via the local directional pattern descriptor applied to two consecutive frames. The Histogram of Oriented Gradients (HOG) descriptor (Dalal and Triggs, 2005) is another widely adopted method due to its invariance to changes in illumination. HOG effectively encapsulates edge information and gradient orientation within an image, making it well-suited for detecting motion in various scenarios. Additionally, robust feature descriptors like the Scale-Invariant Feature Transform (SIFT) (Lowe, 2004) and the Speeded Up Robust Feature (SURF) (Bay et al., 2006) have been incorporated into optical flow frameworks to overcome the problems posed by illumination variation. Chen et al. (2023) introduce a log-correlation transform (LCT) descriptor which operates on feature maps derived from a neighborhood of pixels to eliminates common illumination parameters shared by near pixels, thereby mitigating the impact of complex illumination changes. These descriptors not only show resistance to changes in illumination but also to others alterations such as image scaling, rotation, and affine distortion, making them particularly useful for increase robustness in optical flow methodologies.

**Domain Change**. These methods tackle illumination variations by employing photometric invariant domains. One commonly used technique is Structure-Texture Decomposition, also known as Cartoon Decomposition. It segments the input image into *structure* and *texture* segments (Xu et al., 2012; Aujol et al., 2006). This procedure assists in differentiating between image components that are differentially influenced by illumination changes. The final image is then synthesized as a linear combination of these components, emphasizing the texture component. Typically, additive illumination variations impact the structure more significantly, leaving the texture component less disturbed. However, this technique could potentially lose relevant information, especially when brightness variations are absent. Wulff and Black (2015) use Principal Component Analysis (PCA) to transition the input frame into a smaller subspace represented by the most relevant features. The dense optical flow field is then determined by estimating the position in the PCA subspace that best accounts for the sparse matches. Reformulating the problem within a physics-based framework, can be an effective method to face challenging illumination variations, as proposed by Liao et al. (2023). In this method, illumination variations are modeled using a linear brightness transformation, incorporating a multiplicative component to account for velocity divergence and an additive diffusion component to mitigate the impact of illumination changes. The resulting physics-based optical flow model under varying illumination (PBOFVI) is then solved through a two-phase optimization procedure with a smoothness-sparsity constraint to preserve motion discontinuities. Photometric invariant color spaces can also be employed to manage illumination changes. Resilience during brightness fluctuations is attained by performing a color channel transformation, which transforms input color images into a color space where channels are invariant to illumination shifts. Investigated color spaces include HSV (Mileva et al., 2007), HSI (van de Weijer and Gevers, 2004), normalized RGB (Zickler et al., 2008), and spherical spaces (Van de Weijer et al., 2005). Frequency filters are another method that exploits the frequency domain to render the optical flow computation invariant to lighting changes. Fleet and Jepson (1990) introduces velocity-tuned filters to extract flow components from varying spatial frequencies of the image. These filters can be applied concurrently to the image at different scales and orientations, leading to more robust flow estimation. Laplacian filtering offers a solution to illumination invariance in optical flow estimation (Babaud et al., 1986). It involves computing the Laplacian of the input images, i.e., the second-order derivative in both spatial dimensions. The Laplacian emphasizes areas with swift intensity changes, enhancing edges and dampening low-frequency components related to gradual illumination changes. By focusing on high-frequency image components, the Laplacian filter can minimize the impact of illumination variations on optical flow estimation, making it more robust to lighting changes. However, excessive filtering might result in the loss of valuable information in areas with subtle intensity gradients. Finally, census transform has attracted attention for its application in optical flow estimation (Zabih and Woodfill, 1994). This method generates a bit string representing the relative pixel values in a patch compared to its center pixel. By ignoring absolute intensity values and focusing solely on the neighborhood structure, the Census signature demonstrates resistance to changes in lighting conditions. Despite its effectiveness in outdoor settings and vehicular driving scenarios, incorporating the Census transform into variational optical flow proves complex due to its non-linear nature.

### *3.4. Outliers, discontinuities, and artifacts*

Traditional optical flow estimation frameworks leverage several assumptions to filter out outliers and augment the quality of the computed optical flow map. Outliers, discontinuities, and artifacts pervade two primary areas: the input frames and the estimated optical flow field. The issues they cause can significantly deteriorate the quality of optical flow estimation, making their effective management critical for enhancing the accuracy and reliability of the optical flow estimation.

**Pre-processing and post-processing in Optical Flow Estimation.** When judiciously applied, pre-processing and post-processing can both mitigate noise and enhance frame structures (Sun et al., 2010), bolstering the accuracy of optical flow estimation (Tomasi and Manduchi, 1998; Sun et al., 2014).

In the realm of image pre-processing, the primary objective is to suppress unwanted fluctuations in images. This reduction in variance sets a stable foundation for subsequent algorithmic processes, ensuring they operate under optimal conditions. A representative category of such pre-processing techniques is the Gaussian filter, which has its roots in the seminal works of Horn and Schunck (1981), which is well known for its efficacy in blurring and noise reduction. Another notable technique is the Partial Differential Equation (PDE) filters, as described by Rudin et al. (1992). Differently, non-local filtering focuses on a broader pixel search area. It enhances images by averaging pixels with similar characteristics. This ensures noise reduction without compromising vital details. This concept is well explained in the work by Liu and Freeman (2010). At the forefront of modern approaches, there is the Hybrid GBF and Gaussian filter (HGBGF) proposed by Tu et al. (2014). This method seeks to sidestep the over-smoothing tendencies of Gaussian kernels while preserving the intrinsic structure of input frames within an efficient computational framework. Frames pre-processing can also increase the efficiency of deep learning models. Dobrički et al. (2022a) offer interesting insights. Driven by the ambition to optimize the efficiency of deep learning models, they carried out a comprehensive evaluation of image pre-processing methodologies. Among their foremost recommendations for pre-processing are grayscale representations, directional derivatives, gradient norms, and Gaussian blurring.

Post-processing refining can significantly enhance the precision of optical flow estimation. In fact, not only some apparent outliers can be removed, but also some inaccurate flow components can be corrected (Tu et al., 2014). Common techniques are Median Filtering (Sun et al., 2010) and Bilateral Filter. The advantage of the latter is that rather than focusing solely on noise reduction, it strikes a balance by preserving vital image edges. This dual focus on spatial proximity and pixel value similarity ensures a comprehensive filtering approach (Xiao et al., 2006). Unfortunately, excessive application of post-processing filters could lead to unintentional outcomes, such as the dreaded over-smoothing or the misdirection of flow information, especially in occluded regions (Xiao et al., 2006). Recognizing these potential pitfalls, the scientific community has proposed strategies for meticulous tuning of denoising filter parameters (Sharmin and Brad, 2012). Tu et al. (2014) proposed Combined Post-Filtering (CPF), which leverages flow edges and occlusion detection to refine the intermediate flow field during the estimations, avoiding loose formations. Post-processing flow refining can also be learned and performed by using convolutional layers, as proposed by the LiteFlowNet model (Hui et al., 2018). Embarking on a novel path, this solution incorporates a feature-driven local convolution layer at each pyramid level. Tailored to sync with the encoder, flow estimate, and occlusion probability map, this approach has proven pivotal in addressing challenges like outliers, discontinuities, and artifacts in optical flow estimation. Yet, one must remain cognizant of the potential increase in model complexity this innovation might bring, beckoning further exploration and research.

**Segmentation.** Integrating segmentation techniques into optical flow estimation has been recognized as an effective strategy to maintain the integrity of flow discontinuities (Black and Anandan, 1996). This is significant as it helps to preserve the accurate depiction of object boundaries and motion within a scene, thus leading to more accurate and robust optical flow estimates. Joint estimation-segmentation models, as proposed by Mémin and Pérez (1998) and Brox and Malik (2010), have proved very effective. These models employ a mix of local smoothness and region-wise parametrization to preserve flow field discontinuities. In essence, they segment the image into distinct regions and then apply the optical flow estimation separately to each region, which allows for differing motion patterns across regions and better preservation of discontinuities at the boundaries. Moreover, a multi-scale approach that fuses segmentation and optical flow was introduced by Nir et al. (2008). This approach utilizes global motion patterns to guide segmentation and then refines the optical flow estimation with the resulting segmentation. This method highlights the reciprocal benefits of using segmentation and optical flow in conjunction, where each can help refine the other to improve overall results. Layered representation provides another route for enhancing optical flow computation. A notable development in this area is the localized layer model by Sevilla-Lara et al. (2016), which is particularly useful for semantic image segmentation. This model treats the image as composed of several layers of objects or regions, each with its own motion pattern, which can help provide more nuanced and detailed optical flow estimates. However, while these techniques provide a promising pathway for improved optical flow estimation, they are not without their challenges. For instance, they often involve additional computational complexity and may require careful parameter tuning.

**Prior Motion Statistics**. The integration of natural motion statistics can significantly enhance the performance of various computer vision tasks, including optical flow estimation. A particularly effective strategy involves the extrapolation of motion pattern statistics and their application to refine flow fields (Roth and Black, 2009). This method contributes to reducing uniformity in areas marked by texture and discontinuities, improving the overall quality of the optical flow estimate. Another efficient model that assists in achieving spatial uniformity is the Field-of-Experts (Roth and Black, 2007), which utilizes an interconnected network of local image models to capture the statistical behaviors of natural scenes.

Maintaining temporal consistency is another critical aspect of accurate optical flow estimation. This can be achieved using techniques like Kalman filtering (KF) (Welch and Bishop, 1995), which is a predictive filtering approach that incrementally refines estimates with each new measurement, thus enhancing their accuracy over time. A notable implementation of this approach is the KalmanFlow 2.0 method by Bao et al. (2019). In this model, each pixel's motion flow is treated as a time-variant state vector and is optimally estimated by considering both the measurement and system noise levels. This method leverages the power of compact features derived from pre-trained convolutional neural networks, which provide a rich, context-aware representation of the image for improved estimation.

Using prior knowledge about motion statistics can lead to superior predictions in various real-world scenarios. For example, Leung et al. (2011) applied this concept to medical tracking by incorporating statistical models of cardiac motion into the optical flow framework. Similarly, Chen et al. (2022) introduced a novel framework for complex fluid flow estimation that combines an advanced optical flow method with a joint motion model based on the Helmholtz decomposition theorem. However, it is important to note that using prior motion statistics is challenging. These might include potential inaccuracies in the prior statistics, the computational complexity of some methods, or the need for careful parameter tuning.

### 3.5. Computational constraints

The need for real-time processing in applications such as robotics, autonomous vehicles, and augmented reality has made efficiency in optical flow estimation a critical focus. This is especially important for resource-constrained devices. Factors such as response time, memory usage, and computational efficiency play a crucial role in determining the overall performance and viability of an optical flow method. Traditional variational methods have explored various strategies to reduce the computational demands of their iterative optimization procedures. On the other hand, deep learning models offer a promising alternative with their potential for parallel computation on GPUs, but they can be hindered by large model sizes and complex architectures.

**Pyramidal Warping** Iteratively aligning one image with another using the estimated motion from previous iterations (Brox et al., 2004b; Bruhn et al., 2005) is a classical approach to reduce computational requirements. This method begins with a coarse, low-resolution version of the images to simplify the problem and facilitate a rapid initial estimate. The process subsequently refines the flow field at progressively higher resolutions, effectively reducing the search space and employing a coarse-to-fine refinement strategy. For instance, SPyNet reduces the complexity seen in earlier models like FlowNet (Fischer et al., 2015) by 96% by employing warping at each level of a pyramid structure. Rather than minimizing a classical objective function at each level, SPyNet uses a convolutional network trained from coarse to fine to predict incremental flow changes, thus ensuring displacements remain minimal. DDVM (Saxena et al., 2023) employs a coarse-to-fine approach, initially estimating flow across the entire field of view at a low resolution and subsequently refining these estimates in a patch-wise manner. This strategy mitigates the challenges associated with training high-resolution diffusion models allowing the inference of high resolution optical flow maps.

**Approximated cost volumes.** Cost volumes represent the matching costs between pixels across frames for various displacement hypotheses. While constructing a full cost volume remains computationally prohibitive for real-world applications, several studies have introduced the concept of "partial" cost volume estimation (Xu et al., 2017). This method limits the search range, either at the window level (Hosni et al., 2012) or globally (Barnard, 1989), incorporating explicit smoothness assumptions to manage computational demands effectively.

Deep learning models such as PWCnet (Sun et al., 2018) and FDFlowNet (Kong and Yang, 2020) have integrated warping and partial cost volume estimation to enhance computational efficiency. PWCnet constructs a partial cost volume at each pyramid level and uses a warping layer to link different levels, thereby estimating large displacement flows more effectively. FDFlowNet introduces a novel U-shaped network architecture that integrates multi-scale warping information more efficiently than traditional pyramid structures. LiteFlowNet (Hui et al., 2018) and its successors (Hui et al., 2021; Hui and Loy, 2020) further refine approximated cost volumes approach by employing a cascaded flow inference method, which allows for early error correction and efficient, accurate flow estimation. These networks utilize novel layers to address challenges such as vague boundaries and outliers, improving descriptor matching and enabling efficient short-range matching with sparse cost volumes. The development of optimized optical flow losses, such as the RCELoss (Fan and Cai, 2024), has demonstrated significant potential in training more efficient deep learning models with fewer parameters.

**Table 1**
Comprehensive comparison of optical flow methods on the MPI Sintel Benchmark, sorted by Clean (train) scores.

| Method | MPI-Sintel | | Parameters |
|---|---|---|---|
| | Clean | Final | |
| VideoFlow | 0.991 | 1.649 | 13.5M |
| DDVM | 1.01 | 2.40 | – |
| FlowFormer | 1.16 | 2.09 | 18.2M |
| RAFT | 1.61 | 2.86 | 5.3M |
| LRFlow | 1.98 | 3.81 | 3.00M |
| LiteFlowNet3 | 2.99 | 4.45 | 5.20M |
| VCN-small | 3.26 | 4.73 | 5.20M |
| LiteFlowNet2 | 3.48 | 4.69 | 6.42M |
| FDFLowNet | 3.71 | 5.11 | 5.79M |
| FlowNet2 | 3.96 | 6.02 | 162.4M |
| EpicFlow | 4.12 | 6.29 | – |
| LiteFlowNet | 4.54 | 5.38 | 5.37M |
| FastFlowNet | 4.89 | 6.08 | 1.37M |
| PWC-Net+ | 5.04 | 5.47 | 8.75M |
| PWC-Net-small | 5.05 | 5.32 | 4.08M |
| DeepFlow | 5.38 | 7.21 | – |
| SpyNet | 6.69 | 8.43 | 1.2M |
| FlowNetS | 7.28 | 8.81 | – |
| FlowNetC | 7.42 | 8.43 | – |
| HS | 8.739 | 9.610 | – |

## 4. Evaluation methods and metrics

Accurate error metrics are paramount for precise comparison of optical flow methods. Optical flow estimation's efficacy is typically gauged by the deviations between the estimated optical flow vectors and the ground truth vectors. The two prevalent measures for this are Endpoint Error (EPE) and Angular Error (AE). While traditional metrics have provided foundational assessments, the increasing intricacy and variety of real-world situations demand a more comprehensive evaluation. This section illustrates the metrics that aim to provide insights into the precision and adaptability of optical flow methodologies across different scenarios.

The *Endpoint Error* (EPE) measures the average Euclidean distance between the estimated optical flow vectors and the ground truth vectors for all pixels in the image. Formally:

$$\mathrm{EPE} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(u_i-\hat{u}_i)^2+(v_i-\hat{v}_i)^2}, \tag{6}$$

where $N$ is the total number of pixels in the image, $u_i$ and $v_i$ are the horizontal and vertical components of the ground truth optical flow vector at the $i$th pixel, respectively, and $\hat{u}_i$ and $\hat{v}_i$ are the horizontal and vertical components of the estimated optical flow vector at the $i$th pixel, respectively.

Table 1 compares some key optical flow methods. We can appreciate how EPE can help to evaluate different model performances under the MPI Sintel dataset (Butler et al., 2012). The *Clean* and *Final* columns indicate different render passes used in the dataset.

The Angular Error (AE) quantifies the difference in orientation between the estimated and the ground truth optical flow vectors:

$$\mathrm{AE} = \arccos\left(\frac{f\cdot\hat{f}+1}{\sqrt{(f\cdot f+1)(\hat{f}\cdot\hat{f}+1)}}\right), \tag{7}$$

where $f$ and $\hat{f}$ represent the ground truth and the estimated optical flow vectors, respectively, and $f \cdot \hat{f}$ represents their dot product. Although AE is well-suited for small displacements, it underestimates large motions, making it less suitable for larger vectors. In contrast, EPE is more appropriate for larger vectors and is more frequently employed in contemporary research.

Such traditional metrics calculate the difference between the ground truth and the estimated optical flow. While effective, these evaluation methods offer a generic view and do not consider the specific challenges of calculating optical flow. This limitation has inspired a new class of better evaluation methods under challenging scenarios (Sun et al., 2010). For instance, the MPI Sintel dataset (Butler et al., 2012) uses traditional EPE and several additional metrics to assess model performance under various conditions, particularly focusing on challenging regions of input frames. MPI Sintel proposes the *EPE unmatched* metric, which measures the error in regions visible in only one of two consecutive frames, while other metrics like d0-10, d10-60, and d60-140 measure EPEs within specific pixel distances from the nearest occlusion boundary. Mehl et al. (2023) introduced an advanced expansion to the MPI Sintel evaluation, incorporating new metrics to assess optical flow estimation with rigid, non-rigid, and sky regions.

The evaluation of optical flow algorithms can also employ robust measures like the Weighted Area Under the Curve (WAUC) (Richter et al., 2017). WAUC calculates inlier rates for a range of thresholds and integrates these rates, assigning higher weights to lower-threshold rates. The thresholds and weights used are inversely proportional, allowing for a more nuanced evaluation of the algorithm's performance. The KITTI-FL metrics (Cabon et al., 2020) provide a different approach, deeming a pixel correctly estimated if the flow endpoint error is less than three pixels or within 5% of the ground truth. This approach illustrates how varying the evaluation criteria can expose different strengths and weaknesses in optical flow estimation algorithms.

While the presented evaluation methods focus on minimizing the numerical difference between the predicted flow and the ground truth, exploiting the geometry of the motion can improve the evaluation. Fan and Cai (2024) introduce the Random Epipolar Constraint Loss (RECLoss), including geometric constraints derived from epipolar geometry into the evaluation loss, shifting the emphasis from numerical accuracy to more interpretable geometric consistency.

In conclusion, it is crucial to consider the specific scenario when evaluating model performance. As we will discuss in Section 6, this factor significantly impacts the performance of the model. Ultimately, the definition of a unique framework for model evaluation remains an open question, underlining the importance of multiple metrics in modern benchmarks. As optical flow estimation continues to cater to more intricate and specific application scenarios, the refinement of these metrics and the development of new ones to better capture the nuances of model performance become increasingly crucial.

## 5. Datasets

Optical flow pioneered the establishment of standard datasets in the early days of computer vision, facilitating rigorous, quantitative evaluations of a plethora of algorithms (Barron et al., 1994). As algorithms and estimation techniques evolved, especially with the advent of deep learning models, there emerged a pressing need for richer and more nuanced datasets. These datasets not only aimed to offer improved methodological comparisons but also to emulate the realism of diverse real-world application scenarios.

During the infancy of this evolution, benchmarks were rather elementary, focusing predominantly on basic transformations like translations or rotations. These transformations were applied to either authentic images or synthetic sequences with predefined motions. Yet, as the field expanded, this foundational simplicity gave way to a more intricate representation. Datasets evolved to mirror the chaotic and unpredictable nature of real-world environments. Consequently, contemporary benchmarks adapted, encapsulating scenarios marked by fluctuating lighting, pronounced displacements, abrupt changes, and a host of other complexities inherent to real-world situations.

In this section, we provide an extensive collection of available optical flow datasets. For every dataset discussed in the following paragraphs, there is a comprehensive description of its resolution, an exact breakdown of frames designated for training and evaluation, and an emphasis on its multimodal attributes. These details are crucial as they can inspire other computational pursuits, particularly in training multifunctional models. Moreover, a special focus is given to datasets supplemented with annotations like scene flow (Zhai et al., 2021), disparity maps, occlusion maps, and depth maps, underscoring their enhanced utility. For readers seeking a concise overview, a summary is presented in Table 2.

**General purpose.** This collection of datasets caters to a range of general-purpose computer vision tasks. We first introduce some pioneering classical datasets that drive the progress of classical optical flow methods, and after that, we will discuss some modern datasets. The Yosemite dataset, as presented by Barron et al. (1994), is an early exemplar of optical flow datasets and comprises synthetic images of a mountainous landscape. It is the first optical flow dataset designed to evaluate different methods over naturalistic images, and it is notable for its variation in noise levels and inclusion of ground truth flow fields. Another noteworthy dataset is UCL-Flow (Mac Aodha et al., 2012), which contains real-world grayscale image pairs with known camera motions. Its real-world image data makes it particularly well-suited for evaluating optical flow algorithms in realistic conditions. Middlebury (Baker et al., 2011) stands as a prominent optical flow dataset, containing a blend of synthetic and real-world images complete with ground truth flow fields. This dataset has garnered wide acclaim and serves as a benchmark for optical flow algorithm comparisons. Moving on, Flying Chairs (Fischer et al., 2015) is the first large dataset designed for training deep neural networks. It takes a novel approach by using synthetic image pairs of randomly arranged flying chairs. This unconventional setup is specifically designed to offer ground truth optical flow data for training and evaluation of optical flow algorithms. Flying Chairs2 (Ilg et al., 2018) builds on the structure of the original Flying Chairs dataset by retaining its foundational structure while integrating additional modalities for more comprehensive evaluations. FlyingThings3D (Mayer et al., 2016) diverges from flying chairs to concentrate on intricate 3D shapes in motion. This synthetic dataset includes an assortment of 3D objects and is not only equipped with ground truth data for optical flow but also depth and disparity data. Consequently, it is an invaluable resource for the development and assessment of scene flow algorithms, which necessitate more intricate data on object motion and depth in 3D space. EDEN (Le et al., 2021) is a specialized multi-modal dataset containing outdoor scenes for nature-oriented applications. Finally, SceneNet RGB-D (McCormac et al., 2017) is a large-scale dataset that incorporates synthetic RGB-D images with ground truth optical flow, depth, and object segmentation data. It is versatile, supporting training and evaluation for a variety of computer vision tasks, including optical flow estimation.

**Animation.** Datasets in the animation category are particularly engaging as they tend to encompass imagery from animated films, which are often rich in detail and complexity. A standout dataset in this domain is the MPI-Sintel dataset (Butler et al., 2012), which is constructed from sequences of the open-source 3D animated short film *Sintel*. The realism in the sequences, coupled with the availability of ground truth optical flow data, makes MPI-Sintel a great dataset for optical flow algorithms. Notably, it features sequences with dynamic illumination changes, large displacements, and occlusions, emulating real-world scenarios, which is imperative for assessing the tenacity of optical flow algorithms in real-life complexities. Contrasting MPI-Sintel, the Monkaa dataset (Mayer et al., 2016) offers synthetic sequences extracted from the animated movie *Monkaa*. Although it also provides ground truth optical flow data, the synthetic environment is somewhat

**Table 2**
Reviewed datasets. For each dataset, we summarize the most important aspects and available information.

| | Dataset name | year | Real world | Scene FLow | Disparity map | Occlusion map | Depth map | Segmentation | total frames | Resolution |
|---|---|---|---|---|---|---|---|---|---|---|
| General porpuse | Yosemite | 1994 | – | – | – | – | – | – | 15 | 316 × 252 |
| | UCL-Flow | 2013 | – | – | – | – | – | – | – | – |
| | Middlebury | 2011 | – | – | – | – | – | – | 72 | 640 × 480 |
| | Flying Chairs | 2015 | – | – | – | – | – | – | 22 872 | 512 × 384 |
| | Flying Chairs 2 | 2018 | – | ✓ | – | ✓ | – | – | 22 872 | 512 × 384 |
| | FlyingThings3D | 2016 | – | ✓ | ✓ | – | – | ✓ | 26 066 | 960 × 540 |
| | EDEN | 2020 | – | – | ✓ | – | ✓ | ✓ | 300 000 | 480 × 640 |
| | SceneNet RGB-D | 2017 | – | – | – | ✓ | ✓ | ✓ | 5,000,000 | 320 × 240 |
| Animation | MPI-Sintel | 2012 | – | – | ✓ | – | ✓ | ✓ | 2194 | 1024 × 436 |
| | Monkaa: | 2016 | – | ✓ | ✓ | – | – | ✓ | 8591 | 960 × 541 |
| | Spring | 2023 | – | ✓ | ✓ | – | – | – | 6000 | 4k |
| | Creative Flow+ | 2019 | – | – | – | ✓ | ✓ | ✓ | 134,000 | 1500 × 150 |
| Driving/Urban | KITTI2015 | 2015 | ✓ | ✓ | – | – | ✓ | ✓ | 600 | 1242 × 375 |
| | Virtual KITTI | 2016 | – | ✓ | – | – | ✓ | ✓ | 21,260 | 1242 × 375 |
| | Virtual Kitti2 | 2020 | – | ✓ | – | – | ✓ | ✓ | 21,260 | 1242 × 375 |
| | Driving (scene flow) | 2016 | – | ✓ | ✓ | – | – | ✓ | 4392 | 960 × 542 |
| | HD1k | 2016 | ✓ | – | ✓ | – | – | ✓ | 3563 | 2560 × 1080 |
| | SynWoodScape | 2022 | – | – | – | – | ✓ | ✓ | 80 000 | 1280 × 966 |
| | VIsual PERception (VIPER) | 2017 | – | – | – | – | – | ✓ | 250 000 | 1920 × 1080 |
| | City Scene | 2022 | – | – | – | – | – | – | 3468 | – |
| Challenging Scene | ChairsSDHom | 2017 | – | – | – | – | – | – | – | 512 × 384 |
| | FlowDark | 2020 | – | – | – | – | – | – | 1200 | – |
| | FVR-600 | 2018 | – | – | – | – | – | – | – | 388 × 584 |
| | NUS-100 | 2018 | ✓ | – | – | – | – | – | – | 388 × 584 |
| | GOF (Gyroscope Optical Flow) | 2021 | ✓ | – | – | – | – | – | 2000 | 600 × 800 |
| Human activities | refresh | 2018 | – | ✓ | – | – | ✓ | – | 83 164 | – |
| | humanFLow | 2019 | – | – | – | – | – | – | 146 020 | 256 × 256 |
| | CrowdFlow | 2018 | – | – | – | – | – | – | 3200 | 1280 × 720 |
| | OmniFlow | 2021 | – | – | – | – | – | – | 23.653 | 2,048 × 2,048 |
| Special cameras | Equirectangular FlyingThing | 2022 | – | – | – | – | – | – | 2211 | – |
| | Flow360 | 2022 | – | ✓ | – | – | ✓ | – | 4000 | 512 × 1025 |
| | EVIMO | 2020 | ✓ | – | – | – | ✓ | ✓ | – | – |
| | EVIMO2 | 2021 | ✓ | – | – | – | ✓ | ✓ | 6000 | 640 × 480 and 2080 × 1552 |
| | SPIFT | 2022 | – | – | – | – | – | – | 55 000 | – |
| | PHM | 2022 | – | – | – | – | – | – | 25 100 | – |
| | eventVision-evbench | 2016 | – | – | – | – | ✓ | – | – | – |
| | SlowFlow | 2017 | ✓ | – | – | ✓ | – | – | 2560 × 1440 | – |
| Dataset generators | AutoFlow | 2021 | – | – | – | – | – | – | – | – |
| | Real flow | 2022 | – | – | – | – | – | – | – | – |
| | Arap_flow | 2021 | – | – | – | – | – | – | – | – |

more controlled than the MPI-Sintel dataset. This controlled setting can be crucial for laying a solid foundation for understanding optical flows before delving into more complex and unpredictable environments. The Spring dataset (Mehl et al., 2023) introduces a different flavor by concentrating on high-resolution synthetic image sequences. What sets the Spring dataset apart is the complexity of motion it encompasses. It is purpose-built for evaluating how well optical flow algorithms can perform in areas with high levels of detail and to appraise their proficiency in handling both rigid and non-rigid regions, as well as sky regions. This dataset is particularly advantageous for analyzing the versatility of algorithms in diverse settings. Lastly, the Creative Flow+ dataset (Shugrina et al., 2019) carves a niche by including artistic image sequences along with the usual ground truth optical flow. Its main objective lies in evaluating optical flow algorithms within the framework of non-realistic, artistic imagery. This departure from traditional imagery poses fresh challenges and paves the way for the application of optical flow algorithms in the burgeoning creative industry. In a field that thrives on innovation, Creative Flow+ could play a pivotal role in heralding novel applications and solutions.

**Driving/Urban Environment.** The datasets tailored for driving and urban environments hold critical importance, especially considering the rapid advancements in autonomous vehicle technology. Evaluating and training optical flow algorithms under this setting can substantially contribute to the safety and performance of autonomous systems as they navigate through complex urban landscapes. A highly revered dataset in this context is the KITTI2015 (Menze and Geiger, 2015), which is designed with a specific focus on autonomous driving. It incorporates real-world stereo image pairs along with the corresponding ground truth optical flow data, all captured from a moving vehicle. The incorporation of real-world data renders KITTI2015 invaluable for assessing algorithms in realistic driving conditions. On the synthetic side, Virtual KITTI (Cabon et al., 2020) and Virtual KITTI 2 (Cabon et al., 2020) serve as synthetic equivalents to the KITTI dataset. What is distinctive about these datasets is that they allow evaluations of optical flow algorithms under a plethora of weather conditions, lighting scenarios, and camera viewpoints. This sort of control over environmental variables is instrumental in synthetic datasets as it permits thorough assessments under diverse conditions that might not be feasible with real-world datasets. The driving (scene flow) dataset (Mayer et al., 2016) is another synthetic dataset that focuses on driving scenes. It furnishes ground truth scene flow, depth, and optical flow data. Its goal is to evaluate optical and scene flow algorithms in driving contexts. Having depth information at hand is especially beneficial as it provides insights into how algorithms respond to objects at varying distances. The HD1k dataset (Kondermann et al., 2016) distinguishes itself through its high-resolution imagery. It comprises real-world sequences captured at 1 K resolution and serves as a benchmark for evaluating how optical flow algorithms grapple with high-resolution images. This is vital for detecting minute details in a scene. SynWoodScape dataset (Sekkat et al., 2022) is another synthetic dataset that includes images of densely populated urban settings with ground truth optical flow. It targets the training and evaluation of optical flow algorithms in scenarios characterized by dense urban traffic and constructions, which require high levels of accuracy. Shifting the focus to visual perception, the Visual PERception (VIPER) dataset (Richter et al., 2017) is tailored for training and evaluating visual perception algorithms, including optical flow. It encompasses synthetic image sequences with ground truth data under a variety of environmental conditions, thus allowing for an analysis of algorithm adaptation to different settings. Lastly, the City Scene dataset (Li et al., 2022) is crafted to simulate procedurally generated cityscapes and incorporates over 300 vehicles with collision detection. With its subsets varying in complexity and configuration, City Scene offers a wealth of resources for pre-training, experimentation, and evaluation of algorithms in simulated urban environments. The inclusion of collision detection adds an extra layer of realism and complexity, making it an excellent tool for developing sophisticated navigation and detection algorithms.

**Challenging Scenes.** Addressing challenging scenarios is pivotal for developing robust optical flow algorithms that can operate efficiently under various conditions. Datasets that focus on challenging scenes often target specific problems like untextured regions, low-light conditions, or weather effects and help in creating algorithms that are resilient to such adversities. ChairsSDHom (Fischer et al., 2015) is a synthetic optical flow dataset that specifically targets the challenges posed by untextured regions and small displacements. It features random configurations of flying chairs that are meticulously designed to mimic realistic scenarios. The deliberate avoidance of overfitting to any specific situation makes ChairsSDHom a versatile tool for understanding and overcoming the difficulties inherent in untextured regions. FlowDark (Zheng et al., 2020) takes on the challenge of low-light conditions. It is a low-light optical flow dataset synthesized by simulating noise models on dark, raw images. This allows models to learn optical flow directly from noisy, low-light images. Moreover, FlowDark encompasses a wide gamut of exposure levels for benchmarking purposes, enabling the training of models that are not only precise in low-light conditions but also exhibit superior performance compared to existing methods. FVR-600 and NUS-100 (Li et al., 2018) blend real rain images with synthesized object motions to create a hybrid dataset named FVR-660 with known ground truths. It consists of 660 sequences. In addition, the NUS-100 dataset contains 100 sequences of real rain and real motion, with the ground truth obtained through human annotation. These datasets are especially valuable for training optical flow algorithms to work effectively with rainy conditions, which is a common and challenging weather condition for vision-based systems. Finally, the GOF (Gyroscope Optical Flow) dataset (Li et al., 2021a) is designed to assess optical flow algorithms that leverage gyroscope data. It contains image sequences coupled with gyroscope data and ground truth optical flow. The integration of gyroscope data is significant as it provides additional information regarding camera motion, which can be invaluable in scenarios where visual data alone may be insufficient or unreliable.

**Human Activities.** When it comes to understanding and analyzing human activities through visual data, optical flow algorithms play an important role. Datasets that focus on human activities range from capturing the subtle motions of a single individual to analyzing the complex movements of crowds. These datasets are essential for applications such as action recognition, human–computer interaction, and video surveillance. The Refresh dataset (Lv et al., 2018) contains real-world image sequences along with ground truth optical flow data. It is specially tailored for evaluating optical flow algorithms in video inpainting and frame interpolation tasks. This dataset is particularly beneficial for applications such as video editing and special effects, where the accurate reconstruction of human movements is essential. The HumanFlow dataset (Ranjan et al., 2018, 2020) comprises two distinct datasets containing image sequences of single and multi-human motion. These datasets provide ground truth optical flow, and they are handy for studying and understanding the intricacies of human motion. Whether it is the detailed movement of a single person or the complex interplay of multiple individuals, HumanFlow serves as a rich resource. CrowdFlow (Schröder et al., 2018) specifically targets crowded scenes, containing image sequences of such scenes along with ground truth optical flow. Evaluating optical flow algorithms on challenging crowd motion is crucial for applications like crowd management and surveillance, where understanding the dynamics of large groups of people is key. The OmniFlow dataset (Seidel et al., 2021) introduces a different facet by focusing on household activities. The dataset contains rendered images of various household activities and the corresponding forward and backward optical flow. Such a dataset is particularly valuable in developing smart homes and robotic applications, where understanding and interpreting human activities within a household setting is vital. Overall, these datasets are indispensable for developing optical flow algorithms that are sensitive and responsive to human movements and have diverse applications in many applications ranging from entertainment to security.

**Special Cameras.** When developing new optical flow solutions, it is crucial to take the diverse imaging technologies and camera configurations into account. Specialized cameras, such as panoramic cameras and event-based cameras, capture data in unique ways, and corresponding datasets provide an opportunity to develop specialized algorithms. The EquirectFlyingThings dataset (Li et al., 2022) is inspired by the FlyingChairs and FlyingThings3D datasets but employs panoramic cameras to capture randomly distributed objects from Stanford's ShapeNet dataset. With a higher object count due to the use of panoramic cameras, the dataset includes 300 rendered objects undergoing random translations and rotations. It is ideal for algorithms focused on panoramic imagery and contains sub-datasets with different quantities of image pairs. The Flow360 dataset (Shi et al., 2022) contains 360-degree panoramic image sequences with ground truth optical flow data. This dataset is specially designed for training and evaluating optical flow algorithms on omnidirectional images, and it is particularly useful for applications like virtual reality and 360-degree video processing. EVIMO (Mitrokhin et al., 2019) and its updated version EVIMO2 (Burner et al., 2022) focus on event-based cameras. These datasets contain event camera data along with ground truth optical flow, which captures changes in the scene with high temporal resolution and is ideal for high-speed and low-light scenarios. The EVentVision-evbench (Barranco et al., 2016) aims at evaluating both frame-free and frame-based optical flow methods in visual navigation tasks. It contains synthetic and real scenes captured with various sensors, including an event-based camera (DAVIS) and RGB+Depth sensors. The dataset provides a comprehensive set of data, including images, events, optical flow, 3D camera motion, scene depth, and calibration procedures, facilitating rigorous algorithm evaluation. SlowFlow (Janai et al., 2017) contains slow-motion video sequences along with ground truth optical flow data. It is designed to evaluate optical flow algorithms, specifically on slow-motion videos, which require accurate motion estimation over long time intervals. The SPIkingly Flying Things (SPIFT) and Photo-realistic High-speed Motion (PHM) (Hu et al., 2022) datasets are both tailored for research involving spiking cameras. SPIFT captures high-speed and unpredictable scenarios, making it valuable for testing algorithms in dynamic environments. On the other hand, PHM contains structured high-speed scenes, offering a controlled setting for validating the performance of algorithms with spiking cameras. In conclusion, datasets created with special cameras provide an opportunity to evaluate and develop optical flow algorithms under unique conditions, including panoramic imaging, event-based data capture, slow motion, and more. These datasets are crucial in advancing the optical flow field, particularly in applications requiring specialized imaging technologies.

**Dataset Generators.** Automating data generation can significantly streamline the process of training and evaluating optical flow algorithms. This approach allows for the creation of diverse and accurate synthetic data that closely approximates real-world scenarios, leading to improved generalization and performance. Among several techniques, AutoFlow (Sun et al., 2021) is an automated system for generating optical flow training data. This method employs learnable hyperparameters to control the motion, shape, and appearance of objects within rendered scenes. The generated data can be used to optimize the performance of optical flow models, helping to create algorithms that are more robust and accurate. Arap_flow (Lê et al., 2021) is another automatic data generative method. It extracts and matches objects from video frames, calculates initial constraints, and applies deformation to generate dense optical flow fields. This method allows for the automatic creation of comprehensive optical flow datasets from pre-existing video data. RealFlow (Han et al., 2022) takes a slightly different approach, creating large-scale optical flow datasets from unlabeled realistic videos. It estimates optical flow between video frames and uses these estimates to synthesize new images and corresponding flows. This method effectively transforms unlabeled video data into valuable training data for optical flow algorithms. Finally, Kubric (Greff et al., 2022) is an open-source Python framework that facilitates the generation of high-quality, task-specific datasets. By interfacing with PyBullet and Blender, it can produce photorealistic scenes with rich annotations, significantly streamlining the dataset generation process. In conclusion, automated dataset generators such as AutoFlow, Arap_flow, RealFlow, and Kubric enable more efficient and effective training of optical flow models. By providing a means to rapidly generate diverse and accurate training data, these methods foster the development of more robust and sophisticated optical flow algorithms.

## 6. Discussion

In 2018, Guney et al. (2018) raised the question *is optical flow solved?*. At that time, the advent of deep learning in optical flow estimation had already led to significant improvements in benchmark datasets. For instance, Guney et al. (2018) reported a 92% accuracy in the KITTI benchmark and an average endpoint error of fewer than 5 pixels in the Sintel benchmark.

The relevance of applying optical flow in real-world situations persists, as challenges in these contexts often surpass those represented in current datasets. Factors such as severe lighting changes, complex motion patterns, occlusions, and texture variations can confound even the most sophisticated models. Furthermore, the effectiveness of deep learning models is often linked to the availability of large amounts of data for training, which is not always possible in numerous application scenarios. Furthermore, real-world applications require not only precision but also computational efficiency, real-time performance, and robustness. Thus, it is unsurprising that traditional variational methods are still widely used and often more successful in real-world applications, whereas the application of deep learning models remains challenging.

This reminder of this section is structured as follows: In Section 6.1, we explore how the availability of current data prevents the application of state-of-the-art deep learning models in numerous real-world scenarios. Finally, in Section 6.2, we suggest innovative research directions that can enhance optical flow estimation in real-world applications.

### *6.1. Data availability for real-world applications*

Deep learning for optical flow estimation is an exciting and rapidly evolving field. This growth is largely fueled by the ability of algorithms to effectively learn from large datasets, as shown by Fischer et al. (2015). Unfortunately, one of the biggest limitations in the adoption of deep learning in real-world applications lies in the deficiency of domain-specific datasets. A thorough analysis of optical flow applications shows a considerable gap between its wide-ranging uses and the availability of dedicated datasets. This disconnection poses a major barrier to fully leveraging the power of contemporary optical flow techniques in specialized fields. A clear example can be seen in areas such as fluid dynamics and medical imaging (Trinh and Daul, 2019). For example, in medical endoscopy imaging, optical flow algorithms are expected to tackle challenges and features that are significantly different from those encountered in the context of traditional datasets. These include the need to handle surfaces that lack discernible features and are laden with artifacts, the need to identify correspondences in complex anatomical structures, and the need to conquer the challenges posed by ground truth depth estimation (Yang et al., 2021).

The scarcity of domain-specific datasets has driven researchers to explore alternative methods, often with suboptimal results. Custom dataset creation is complex and limited (Mendes et al., 2022), while synthetic data, though invaluable, often fails to capture real-world photometric details, leading to a disconnect between synthetic and real-world data distributions (Butler et al., 2012; Mayer et al., 2016). Benchmark datasets like MPI-Sintel and KITTI are crucial for evaluating optical flow models, but high performance on these benchmarks does not guarantee generalization to other domains. Unsupervised models, trained on abundant unlabeled video data, offer a solution by avoiding training-test data mismatches. These models are training by regularizing prediction with domain knowledge assumptions like photometric consistency, or smooth motion consistency (Dobrički et al., 2022b; Jonschkowski et al., 2020). Representative model of this category are DDFlow (Liu et al., 2019), which introduce self-supervised learning to distil reliable predictions from a teacher network to guide a student network in learning optical flow from unlabeled data and OAFlow (Li et al., 2021b) uses photometric loss between the target image and the warped subsequent image, while incorporating an occlusion map to improve accuracy by focusing the loss on non-occluded regions. However, despite unsupervised methods are a promising direction, unsupervised models still lag behind supervised ones (Jonschkowski et al., 2020; Yin et al., 2022; Yang et al., 2021; Mocanu et al., 2021), leaving room for improvements.

### *6.2. Future research directions*

The field of optical flow estimation is on the brink of transformative innovations that could overcome existing limitations and expand the range of possibilities. As such, it opens the way to several promising research directions.

Unsupervised learning has the potential to revolutionize optical flow estimation by eliminating the need for labeled data. However, a drawback is its generally lower accuracy compared to supervised learning. Jonschkowski et al. (2020) demonstrate how in-depth model investigation can lead to significant improvements, focusing on optimizing photometric losses, occlusion estimation, self-supervision, smoothness constraints, and other factors like pretraining, image resolution, and data augmentation. While photometric losses have been explored for training unsupervised methods, epipolar constraint loss functions like RECLoss (Fan and Cai, 2024) could further enhance these methods with more robust domain knowledge and reduce reliance on synthetic data.

Semi-supervised learning could offer a balanced solution, utilizing labeled data where possible and supplementing it with insights gained through unsupervised learning. This combination could lead to the development of optical flow models that are both proficient and free from data availability constraints (Yin et al., 2022).

As we develop more sophisticated models, the datasets that feed them must also evolve. Automated dataset creation represents a future where algorithms and computer-generated environments collaborate to create custom datasets (Mayer et al., 2016). This approach brings two significant benefits: it reduces resource demands and accelerates the pace at which models can be improved. Crucially, it opens up the possibility of creating datasets specific to niche application areas, laying the groundwork for specialized optical flow models. Regardless of how advanced our algorithms become, they cannot make up for a lack of high-quality, realistic datasets. Thus, innovating data collection techniques is essential. High frame rate cameras could be useful to create new datasets with high realism (Janai et al., 2017) thanks to their ability to capture detailed motion. These realistic datasets could ensure that our models, while theoretically sound, can handle the unpredictability and variety of real-world scenarios.

The Spring dataset (Mehl et al., 2023) provides a glimpse into the evolution of benchmarks. In comparison to its predecessors like MPI-Sintel, the Spring dataset distinguishes itself with its high resolution, extensive range of scenes, and complex motion types. Additionally, its highly detailed ground truth and its versatility for various tasks such as scene flow, optical flow, and stereo make it an invaluable resource for researchers. These realistic datasets could ensure the development of models that can handle the unpredictability and variety of real-world scenarios. Moreover, as models face diverse challenges like large displacements, changes in lighting, and occlusions, the metrics used to assess their performance need to adapt as well. The commonly used End-Point-Error (EPE) loss tends to have a narrow focus in its evaluation, often ignoring the relationship between image pairs and optical flow (Savian et al., 2020). Transitioning towards more comprehensive metrics that can holistically assess model performance against a range of challenges is essential.

The way optical flow models are trained is just as important as their architecture. Research has shown that even slight modifications to training strategies can lead to significant improvements. For example, reordering training data or adjusting training schedules can boost performance by 20%–30% (Ilg et al., 2017), while including geometric constraints on the loss terms can significantly improve performance (Fan and Cai, 2024). This underlines the need to critically examine and innovate upon the training methodologies in use. Encouraging experimentation and analysis could reveal new strategies that maximize performance without requiring more complex model architectures. Tools that automate certain aspects of model development can play a key role in enhancing performance. Autoflow (Sun et al., 2021), for instance, shows how automated tools can be used to fine-tune popular deep learning models like PWC-Net and RAFT. By automating repetitive and time-consuming tasks, researchers can concentrate on more fundamental aspects of model development.

Lastly, diffusion models (Ho et al., 2020, 2022) have emerged as a promising direction in computer vision (Croitoru et al., 2023), departing from traditional approaches and offering a unique perspective by modeling the gradual transformation of data from noise to structured images. This process is governed by a Markov chain of diffusion steps, where each step progressively refines the image representation by denoising the input. By learning to reverse this diffusion process, the model gains the remarkable ability to generate high-quality images from random noise. This shift in paradigm has yielded impressive results in various domains, from generating photorealistic images of faces and landscapes to synthesizing creative artwork. Saxena et al. (2023) leverage diffusion processes, to enhances optical flow estimation by capturing temporal consistency across image sequences. Its effectiveness is demonstrated by its outperformance of traditional baselines on various benchmarks, indicating both improved accuracy and robustness in diverse conditions. Future research can extend this work by integrating the model into broader video processing frameworks, potentially leading to advancements in video analysis, compression, and enhancement.

## 7. Conclusion

Optical flow estimation has come a long way since its introduction. Advances in deep learning have substantially improved its accuracy and efficiency. However, current benchmarks and measurements limit further progress and do not adequately represent real-world scenarios. To advance the field, future research should focus on developing more representative standards, refining evaluation metrics, and improving pipeline design. By striking a balance between accuracy and resource constraints, incorporating domain knowledge, and optimizing training procedures, the computer vision community can drive the development of more robust and efficient optical flow methods for real-world applications.

## CRediT authorship contribution statement

**Andrea Alfarano:** Writing – original draft, Investigation, Formal analysis, Conceptualization. **Luca Maiano:** Writing – review & editing. **Lorenzo Papa:** Writing – review & editing. **Irene Amerini:** Project administration, Methodology.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

No data was used for the research described in the article.

## Acknowledgments

This study has been partially supported by SERICS (PE00000014) under the MUR National Recovery and Resilience Plan funded by the European Union - NextGenerationEU and Sapienza University of Rome project "EV2" (003_009_22).

## References

Agarwal, S., Farid, H., Gu, Y., 2020. Detecting deep-fake videos from appearance and behavior. In: 2020 IEEE International Workshop on Information Forensics and Security. WIFS, IEEE, pp. 1–6.

Agarwal, A., Gupta, S., Singh, D.K., 2016. Review of optical flow technique for moving object detection. In: 2016 2nd International Conference on Contemporary Computing and Informatics. IC3I, IEEE, pp. 409–413.

Amerini, I., Galteri, L., Caldelli, R., Del Bimbo, A., 2019. Deepfake video detection through optical flow based cnn. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops.

Andrade-Miranda, G., Bernardoni, N.H., Godino-Llorente, J.I., Cruz, H., 2018. Vocal folds dynamics by means of optical flow techniques: A review of the methods. Adv. Signal Process. Rev..

Angelini, E.D., Gerard, O., 2006. Review of myocardial motion estimation methods from optical flow tracking on ultrasound data. In: 2006 International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 1537–1540.

Anthwal, S., Ganotra, D., 2019. An overview of optical flow-based approaches for motion segmentation. J. Imaging Sci. 67 (5), 284–294. http://dx.doi.org/10.1080/13682199.2019.1641316.

Aujol, J.-F., Gilboa, G., Chan, T., Osher, S., 2006. Structure-texture image decomposition—modeling, algorithms, and parameter selection. Int. J. Comput. Vis. 67, 111–136.

Babaud, J., Witkin, A.P., Baudin, M., Duda, R.O., 1986. Laplacian pyramid as a compact image code. In: Readings in Computer Vision. Elsevier, pp. 671–679.

Bailer, C., Taetz, B., Stricker, D., 2017. A benchmark dataset and evaluation methodology for video object segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 724–732.

Bailer, C., Taetz, B., Stricker, D., 2019. Flow fields: Dense correspondence fields for highly accurate large displacement optical flow estimation. IEEE Trans. Pattern Anal. Mach. Intell. 41 (8), 1879–1892. http://dx.doi.org/10.1109/TPAMI.2018.2859970.

Baker, S., Scharstein, D., Lewis, J.P., Roth, S., Black, M.J., Szeliski, R., 2011. A database and evaluation methodology for optical flow. Int. J. Comput. Vis. 92 (1), 1–31. http://dx.doi.org/10.1007/s11263-010-0390-2.

Bao, W., Zhang, X., Chen, L., Gao, Z., 2019. KalmanFlow 2.0: Efficient video optical flow estimation via context-aware kalman filtering. IEEE Trans. Image Process. 28 (9), 4233–4246.

Barnard, S.T., 1989. Stochastic stereo matching over scale. Int. J. Comput. Vis. 3 (1), 17–32.

Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.B., 2009. PatchMatch: A randomized correspondence algorithm for structural image editing. 28, (3), ACM, p. 24,

Barranco, F., Fermuller, C., Aloimonos, Y., Delbruck, T., 2016. A dataset for visual navigation with neuromorphic methods. Front. Neurosci. 10, 49.

Barron, J.L., Fleet, D.J., Beauchemin, S.S., 1994. Performance of optical flow techniques. Int. J. Comput. Vis. 12, 43–77.

Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2006. SURF: Speeded up robust features. In: Computer vision–ECCV 2006. Springer, pp. 404–417.

Benameur, N., Arous, Y., Abdallah, N.B., Kraiem, T., 2017. The assessment of left ventricular function in MRI using the detection of myocardial borders and optical flow approaches: A review. Int. J. Cardiovasc. Pract. 2 (4), 73–75.

Black, M.J., Anandan, P., 1996. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. Comput. Vis. Image Underst. 63 (1), 75–104.

Brox, T., Bruhn, A., Papenberg, N., Weickert, J., 2004a. High accuracy optical flow estimation based on a theory for warping. In: European conference on computer vision. Springer, pp. 25–36.

Brox, T., Bruhn, A., Papenberg, N., Weickert, J., 2004b. High accuracy optical flow estimation based on a theory for warping. In: Pajdla, T., Matas, J. (Eds.), Computer Vision - ECCV 2004. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 25–36.

Brox, T., Malik, J., 2010. Large displacement optical flow: descriptor matching in variational motion estimation. IEEE Trans. Pattern Anal. Mach. Intell. 33 (3), 500–513.

Bruhn, A., Weickert, J., Schnörr, C., 2005. Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. Int. J. Comput. Vis. 61, 211–231.

Burner, L., Mitrokhin, A., Fermüller, C., Aloimonos, Y., 2022. EVIMO2: An event camera dataset for motion segmentation, optical flow, structure from motion, and visual inertial odometry in indoor scenes with monocular or stereo algorithms. arXiv preprint arXiv:2205.03467.

Butler, D.J., Wulff, J., Stanley, G.B., Black, M.J., 2012. A naturalistic open source movie for optical flow evaluation. In: Fitzgibbon, A., et al. (Eds.), European Conf. on Computer Vision. ECCV, In: Part IV, LNCS 7577, Springer-Verlag, pp. 611–625.

Cabon, Y., Murray, N., Humenberger, M., 2020. Virtual KITTI 2. http://dx.doi.org/10.48550/ARXIV.2001.10773, arXiv URL https://arxiv.org/abs/2001.10773.

Caldelli, R., Galteri, L., Amerini, I., Del Bimbo, A., 2021. Optical flow based CNN for detection of unlearnt deepfake manipulations. Pattern Recognit. Lett. 146, 31–37.

Chao, H., Gu, Y., Napolitano, M., 2013. A survey of optical flow techniques for UAV navigation applications. In: 2013 International Conference on Unmanned Aircraft Systems. ICUAS, pp. 710–716. http://dx.doi.org/10.1109/ICUAS.2013.6564752.

Chao, H., Gu, Y., Napolitano, M., 2014. A survey of optical flow techniques for robotics navigation applications. J. Intell. Robot. Syst. 73 (1), 361–372.

Chen, J., Duan, H., Song, Y., Cai, Z., Yang, G., 2023. Optical flow computation for video under the dynamic illumination. IEEE Trans. Multimed. 25, 6285–6300. http://dx.doi.org/10.1109/TMM.2022.3207583.

Chen, J., Duan, H., Song, Y., Cai, Z., Yang, G., Liu, T., 2022. Motion estimation for complex fluid flows using helmholtz decomposition. IEEE Trans. Circuits Syst. Video Technol..

Chen, Z., Jin, H., Lin, Z., Cohen, S., Wu, Y., 2013. Large displacement optical flow from nearest neighbor fields. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2443–2450.

Chen, Q., Koltun, V., 2016. Full flow: Optical flow estimation by global optimization over regular grids. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4706–4714.

Chen, K., Lorenz, D.A., 2012. Image sequence interpolation based on optical flow, segmentation, and optimal control. IEEE Trans. Image Process. 21 (3), 1020–1030. http://dx.doi.org/10.1109/TIP.2011.2179305.

Croitoru, F.-A., Hondru, V., Ionescu, R.T., Shah, M., 2023. Diffusion models in vision: A survey. IEEE Trans. Pattern Anal. Mach. Intell..

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1. CVPR'05, IEEE, pp. 886–893.

van de Weijer, J., Gevers, T., 2004. Robust optical flow from photometric invariants. In: 2004 International Conference on Image Processing, 2004, Vol. 3. ICIP '04, pp. 1835–1838 Vol. 3. http://dx.doi.org/10.1109/ICIP.2004.1421433.

Dinesh Reddy, N., Vo, M., Narasimhan, S.G., 2019. Occlusion-net: 2D/3D occluded keypoint localization using graph networks. In: The IEEE Conference on Computer Vision and Pattern Recognition. CVPR.

Dobrički, T., Oh, Y., Ko, H., Kim, T., Kim, D., Hong, B.-W., 2022a. Comparison of optical flow image preprocessing options for state of the art deep learning models. In: 2022 13th International Conference on Information and Communication Technology Convergence. ICTC, IEEE, pp. 585–587.

Dobrički, T., Zhuang, X., Won, K.J., Hong, B.-W., 2022b. Survey on unsupervised learning methods for optical flow estimation. In: 2022 13th International Conference on Information and Communication Technology Convergence. ICTC, IEEE, pp. 591–594.

Editya, A.S., Ahmad, T., Studiawan, H., 2022. Direction estimation of drone collision using optical flow for forensic investigation. In: 2022 10th International Symposium on Digital Forensics and Security. ISDFS, IEEE, pp. 1–6.

Fan, Z., Cai, Z., 2024. Random epipolar constraint loss functions for supervised optical flow estimation. Pattern Recognit. 148, 110141. http://dx.doi.org/10.1016/j.patcog.2023.110141, URL https://www.sciencedirect.com/science/article/pii/S0031320323008385.

Fassold, H., 2022. A qualitative investigation of optical flow algorithms for video denoising. In: Proceedings of the Future Technologies Conference, Vol. 2. FTC 2022, Springer, pp. 765–775.

Fischer, P., Dosovitskiy, A., Ilg, E., Häusser, P., Hazırbaş, C., Golkov, V., Van der Smagt, P., Cremers, D., Brox, T., 2015. Flownet: Learning optical flow with convolutional networks. arXiv preprint arXiv:1504.06852.

Fleet, D.J., Jepson, A.D., 1990. Computation of component image velocity from local phase information. Int. J. Comput. Vis. 5, 77–104.

Fortun, D., Bouthemy, P., Kervrann, C., 2015. Optical flow modeling and computation: A survey. Comput. Vis. Image Underst. 134, 1–21. http://dx.doi.org/10.1016/j.cviu.2015.02.008, Image Understanding for Real-world Distributed Video Networks, URL https://www.sciencedirect.com/science/article/pii/S1077314215000429.

Geiger, A., Lenz, P., Stiller, C., Urtasun, R., 2012. Are we ready for autonomous driving? The kitti vision benchmark suite. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 3354–3361.

Ghalenoei, E., Sharifi, M., Hasanlou, M., 2014. Investigation of optical flow techniques for extracting non-rigid sea surface currents. Int. Arch. Photogramm. Remote Sens. 40 (2), 121.

Gibson, J.J., 1950. The Perception of the Visual World. Houghton Mifflin.

Gkountakos, K., Ioannidis, K., Tsikrika, T., Vrochidis, S., Kompatsiaris, I., 2020. A crowd analysis framework for detecting violence scenes. In: Proceedings of the 2020 International Conference on Multimedia Retrieval. ICMR '20, Association for Computing Machinery, New York, NY, USA, pp. 276–280. http://dx.doi.org/10.1145/3372278.3390725.

Greff, K., Belletti, F., Beyer, L., Doersch, C., Du, Y., Duckworth, D., Fleet, D.J., Gnanapragasam, D., Golemo, F., Herrmann, C., et al., 2022. Kubric: A scalable dataset generator. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3749–3761.

Guney, F., Sevilla-Lara, L., Sun, D., Wulff, J., 2018. " What is optical flow for?": Workshop results and summary. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops.

Han, Y., Luo, K., Luo, A., Liu, J., Fan, H., Luo, G., Liu, S., 2022. RealFlow: EM-based realistic optical flow dataset generation from videos. In: Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIX. Springer, pp. 288–305.

Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. Adv. Neural Inf. Process. Syst. 33, 6840–6851.

Ho, J., Salimans, T., Gritsenko, A., Chan, W., Norouzi, M., Fleet, D.J., 2022. Video diffusion models. arXiv:2204.03458.

Horn, B.K., Schunck, B.G., 1981. Determining optical flow. Artificial Intelligence 17 (1), 185–203. http://dx.doi.org/10.1016/0004-3702(81)90024-2, URL https://www.sciencedirect.com/science/article/pii/0004370281900242.

Hosni, A., Rhemann, C., Bleyer, M., Rother, C., Gelautz, M., 2012. Fast cost-volume filtering for visual correspondence and beyond. IEEE Trans. Pattern Anal. Mach. Intell. 35 (2), 504–511.

Hu, L., Zhao, R., Ding, Z., Ma, L., Shi, B., Xiong, R., Huang, T., 2022. Optical flow estimation for spiking camera. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 17844–17853.

Huang, J.-F., Chen, S.-L., 2014. Detection of violent crowd behavior based on statistical characteristics of the optical flow. In: 2014 11th International Conference on Fuzzy Systems and Knowledge Discovery. FSKD, pp. 565–569. http://dx.doi.org/10.1109/FSKD.2014.6980896.

Huang, X., Fernandez-Rojas, R., Ou, K.-L., 2016. Cortical activation investigation by optical flow and wavelet analysis using near-infrared spectroscopy. In: 2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics. BioRob, IEEE, pp. 1307–1312.

Huang, Z., Shi, X., Zhang, C., Wang, Q., Cheung, K.C., Qin, H., Dai, J., Li, H., 2022. Flowformer: A transformer architecture for optical flow. In: Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVII. Springer, pp. 668–685.

Hui, T.-W., Loy, C.C., 2020. Liteflownet3: Resolving correspondence ambiguity for more accurate optical flow estimation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16. Springer, pp. 169–184.

Hui, T.-W., Tang, X., Loy, C.C., 2018. Liteflownet: A lightweight convolutional neural network for optical flow estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 8981–8989.

Hui, T.-W., Tang, X., Loy, C.C., 2021. A lightweight optical flow CNN —Revisiting data fidelity and regularization. IEEE Trans. Pattern Anal. Mach. Intell. 43 (8), 2555–2569. http://dx.doi.org/10.1109/TPAMI.2020.2976928.

Husseini, S., 2017. A survey of optical flow techniques for object tracking.

Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., Brox, T., 2017. Flownet 2.0: Evolution of optical flow estimation with deep networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2462–2470.

Ilg, E., Saikia, T., Keuper, M., Brox, T., 2018. Occlusions, motion and depth boundaries with a generic network for disparity, optical flow or scene flow estimation. In: European Conference on Computer Vision. ECCV, URL http://lmb.informatik.uni-freiburg.de/Publications/2018/ISKB18.

Ince, S., Konrad, J., 2008. Occlusion-aware optical flow estimation. IEEE Trans. Image Process. 17 (8), 1443–1451. http://dx.doi.org/10.1109/TIP.2008.925381.

Janai, J., Guney, F., Ranjan, A., Black, M.J., Geiger, A., 2018. Unsupervised learning of multi-frame optical flow with occlusions. In: European Conference on Computer Vision. ECCV.

Janai, J., Guney, F., Wulff, J., Black, M.J., Geiger, A., 2017. Slow flow: Exploiting high-speed cameras for accurate and diverse optical flow reference data. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3597–3607.

Jeong, J., Lin, J.M., Porikli, F., Kwak, N., 2022. Imposing consistency for optical flow estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3181–3191.

Jiang, L., Li, R., Wu, W., 2019. Deepfake video detection using recurrent neural networks. In: 2019 15th IEEE International Conference on Advanced Video and Signal Based Surveillance. AVSS, IEEE, pp. 1–6.

Jonschkowski, R., Stone, A., Barron, J.T., Gordon, A., Konolige, K., Angelova, A., 2020. What matters in unsupervised optical flow. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16. Springer, pp. 557–572.

Kajo, I., Malik, A.S., Kamel, N., 2015. Motion estimation of crowd flow using optical flow techniques: A review. In: 2015 9th International Conference on Signal Processing and Communication Systems. ICSPCS, IEEE, pp. 1–9.

Kondermann, D., Nair, R., Honauer, K., Krispin, K., Andrulis, J., Brock, A., Gussefeld, B., Rahimimoghaddam, M., Hofmann, S., Brenner, C., et al., 2016. The hci benchmark suite: Stereo and flow ground truth with uncertainties for urban autonomous driving. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 19–28.

Kong, L., Yang, J., 2020. Fdflownet: Fast optical flow estimation using a deep lightweight network. In: 2020 IEEE International Conference on Image Processing. ICIP, IEEE, pp. 1501–1505.

Le, H.-A., Mensink, T., Das, P., Karaoglu, S., Gevers, T., 2021. Eden: Multimodal synthetic dataset of enclosed garden scenes. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 1579–1589.

Lê, H.-Â., Nimbhorkar, T., Mensink, T., Baslamisli, A.S., Karaoglu, S., Gevers, T., 2021. Automatic generation of dense non-rigid optical flow. Comput. Vis. Image Underst. 212, 103274.

Leung, K.E., Danilouchkine, M.G., van Stralen, M., de Jong, N., van der Steen, A.F., Bosch, J.G., 2011. Left ventricular border tracking using cardiac motion models and optical flow. Ultrasound Med. Biol. 37 (4), 605–616.

Li, Y., Barnes, C., Huang, K., Zhang, F.L., 2022. Deep 360° optical flow estimation based on multi-projection fusion. http://dx.doi.org/10.48550/ARXIV.2208.00776, arXiv URL https://arxiv.org/abs/2208.00776.

Li, H., Luo, K., Liu, S., 2021a. GyroFlow: Gyroscope-guided unsupervised optical flow learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12869–12878.

Li, Y., Lyu, S., 2018. Detection of deep network generated fake images using disparities in color components. In: International Workshop on Digital Watermarking. Springer, pp. 147–157.

Li, R., Tan, R.T., Cheong, L.-F., 2018. Robust optical flow in rainy scenes. In: Proceedings of the European Conference on Computer Vision. ECCV.

Li, J., Zhao, J., Song, S., Feng, T., 2021b. Occlusion aware unsupervised learning of optical flow from video. In: Osten, W., Nikolaev, D.P., Zhou, J. (Eds.), Thirteenth International Conference on Machine Vision, Vol. 11605. SPIE, International Society for Optics and Photonics, pp. 224–231. http://dx.doi.org/10.1117/12.2588381.

Liao, X., Cai, Z., Chen, J., Liu, T., huang Lai, J., 2023. Physics-based optical flow estimation under varying illumination conditions. Signal Process., Image Commun. 117, 117007. http://dx.doi.org/10.1016/j.image.2023.117007, URL https://www.sciencedirect.com/science/article/pii/S0923596523000899.

Lin, F., Fookes, C., Chandran, V., Sridharan, S., 2005. Investigation into optical flow super-resolution for surveillance applications. In: WDIC 2005: APRS Workshop on Digital Image Computing: Workshop Proceedings. University of QLD, pp. 73–78.

Liu, C., Freeman, W.T., 2010. A high-quality video denoising algorithm based on reliable motion estimation. In: Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part III 11. Springer, pp. 706–719.

Liu, P., King, I., Lyu, M.R., Xu, J., 2019. Ddflow: Learning optical flow with unlabeled data distillation. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, No. 01. pp. 8770–8777.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60 (2), 91–110.

Lucas, B.D., Kanade, T., 1981. An iterative image registration technique with an application to stereo vision. In: IJCAI'81: 7th International Joint Conference on Artificial Intelligence, Vol. 2. pp. 674–679.

Lv, Z., Kim, K., Troccoli, A., Rehg, J., Kautz, J., 2018. Learning rigidity in dynamic scenes with a moving camera for 3D motion field estimation. In: ECCV.

Mac Aodha, O., Humayun, A., Pollefeys, M., Brostow, G.J., 2012. Learning a confidence measure for optical flow. IEEE Trans. Pattern Anal. Mach. Intell. 35 (5), 1107–1120.

Makansi, O., Ilg, E., Brox, T., 2017. End-to-end learning of video super-resolution with motion compensation. http://dx.doi.org/10.48550/ARXIV.1707.00471, arXiv URL https://arxiv.org/abs/1707.00471.

Matern, F., Riess, C., Stamm, M.C., 2019. Exploiting visual artifacts to expose deepfakes and face manipulations. In: 2019 IEEE Winter Applications of Computer Vision Workshops. WACVW, IEEE, pp. 83–92.

Mathur, R., 2020. Evaluation datasets and benchmarks for optical flow algorithms: A review.

Mayer, N., Ilg, E., Hausser, P., Fischer, P., Cremers, D., Dosovitskiy, A., Brox, T., 2016. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4040–4048.

McCormac, J., Handa, A., Leutenegger, S., Davison, A.J., 2017. Scenenet rgb-d: Can 5 m synthetic images beat generic imagenet pre-training on indoor segmentation?. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2678–2687.

Mehl, L., Schmalfuss, J., Jahedi, A., Nalivayko, Y., Bruhn, A., 2023. Spring: A high-resolution high-detail dataset and benchmark for scene flow, optical flow and stereo arXiv preprint arXiv:2303.01943.

Mémin, E., Pérez, P., 1998. Dense estimation and object-based segmentation of the optical flow with robust techniques IEEE Trans. Image Process. 7 (5), 703–719.

Mendes, L.P., Ricardo, A.M., Bernardino, A.J., Ferreira, R.M., 2022. A comparative study of optical flow methods for fluid mechanics Exp. Fluids 63 (1), 7.

Menze, M., Geiger, A., 2015. Object scene flow for autonomous vehicles In: Conference on Computer Vision and Pattern Recognition. CVPR.

Mileva, Y., Bruhn, A., Weickert, J., 2007. Illumination-robust variational optical flow with photometric invariants In: Pattern Recognition: 29th DAGM Symposium, Heidelberg, Germany, September 12-14, 2007. Proceedings 29. Springer, pp. 152–162.

Mitrokhin, A., Ye, C., Fermüller, C., Aloimonos, Y., Delbruck, T., 2019. EV-IMO: Motion segmentation dataset and learning pipeline for event cameras In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, IEEE, pp. 6105–6112.

Mocanu, S., Moody, A.R., Khademi, A., 2021. Flowreg: Fast deformable unsupervised medical image registration using optical flow arXiv preprint arXiv:2101.09639.

Mohamed, M.A., Rashwan, H.A., Mertsching, B., García, M.A., Puig, D., 2014. Illumination-robust optical flow using a local directional pattern IEEE Trans. Circuits Syst. Video Technol. 24 (9), 1499–1508.

Nir, T., Bruckstein, A.M., Kimmel, R., 2008. Over-parameterized variational optical flow Int. J. Comput. Vis. 76, 205–216.

Osman, A., Ovinis, M., 2020. A review of in-situ optical flow measurement techniques in the Deepwater Horizon oil spill Measurement 153, 107396.

Osman, A., Ovinis, M., Osei, H., Hashim, F., 2016. A review of optical flow models applied for fluid motion estimation ARPN J. Eng. Appl. Sci. 11 (20), 12181–12186.

Radhakrishnan, R., Sharma, D., Murthy, V., 2017. A review on particle image velocimetry and optical flow methods in riverine environment

Ranjan, A., Black, M.J., 2017. Optical flow estimation using a spatial pyramid network In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4161–4170.

Ranjan, A., Hoffmann, D.T., Tzionas, D., Tang, S., Romero, J., Black, M.J., 2020. Learning multi-human optical flow Int. J. Comput. Vis. 128 (4), 873–890. http://dx.doi.org/10.1007/s11263-019-01279-w.

Ranjan, A., Romero, J., Black, M.J., 2018. Learning human optical flow http://dx.doi.org/10.48550/ARXIV.1806.05666, arXiv URL https://arxiv.org/abs/1806.05666.

Raudies, F., Neumann, H., 2012. A review and evaluation of methods estimating ego-motion Comput. Vis. Image Underst. 116 (5), 606–633. http://dx.doi.org/10.1016/j.cviu.2011.04.004, URL https://www.sciencedirect.com/science/article/pii/S1077314212000021.

Ren, J., Reyes, N., Barczak, A., Scogings, C., Liu, M., 2018. An investigation of skeleton-based optical flow-guided features for 3D action recognition using a multi-stream CNN model In: 2018 IEEE 3rd International Conference on Image, Vision and Computing. ICIVC, IEEE, pp. 199–203.

Revaud, J., Weinzaepfel, P., Harchaoui, Z., Schmid, C., 2015. Epicflow: Edge-preserving interpolation of correspondences for optical flow In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1164–1172.

Richter, S.R., Hayder, Z., Koltun, V., 2017. Playing for benchmarks In: Proceedings of the IEEE International Conference on Computer Vision. ICCV.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, pp. 234–241.

Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., Nießner, M., 2019. FaceForensics++: Learning to detect manipulated facial images In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1–11.

Roth, S., Black, M.J., 2007. On the spatial statistics of optical flow Int. J. Comput. Vis. 74, 33–50.

Roth, S., Black, M.J., 2009. Field of experts Int. J. Comput. Vis. 82 (2), 205–229.

Rudin, L.I., Osher, S., Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms Physica D 60 (1–4), 259–268.

Sabir, E., Cheng, J., Jaiswal, A., AbdAlmageed, W., 2019. Deepfake detection by analyzing convolutional traces

Savian, S., Elahi, M., Tillo, T., 2020. Optical flow estimation with deep learning, a survey on recent advances In: Jiang, R., Li, C.-T., Crookes, D., Meng, W., Rosenberger, C. (Eds.), Deep Biometrics. Springer International Publishing, Cham, pp. 257–287. http://dx.doi.org/10.1007/978-3-030-32583-1_12.

Saxena, S., Herrmann, C., Hur, J., Kar, A., Norouzi, M., Sun, D., Fleet, D.J., 2023. The surprising effectiveness of diffusion models for optical flow and monocular depth estimation In: Thirty-Seventh Conference on Neural Information Processing Systems. URL https://openreview.net/forum?id=jDIlzSU8wJ.

Schröder, G., Senst, T., Bochinski, E., Sikora, T., 2018. Optical flow dataset and benchmark for visual crowd analysis In: 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance. AVSS, pp. 1–6. http://dx.doi.org/10.1109/AVSS.2018.8639113.

Seidel, R., Apitzsch, A., Hirtz, G., 2021. OmniFlow: Human omnidirectional optical flow In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 3678–3681.

Sekkat, A.R., Dupuis, Y., Kumar, V.R., Rashed, H., Yogamani, S., Vasseur, P., Honeine, P., 2022. SynWoodScape: Synthetic surround-view fisheye camera dataset for autonomous driving IEEE Robot. Autom. Lett. 7 (3), 8502–8509.

Sevilla-Lara, L., Sun, D., Jampani, V., Black, M.J., 2016. Optical flow with semantic segmentation and localized layers In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3889–3898.

Shah, Tafseer, S., Haider, 2021. Traditional and modern strategies for optical flow: an investigation In: SN Applied Sciences. http://dx.doi.org/10.1007/s42452-021-04227-x.

Sharmin, N., Brad, R., 2012. Optimal filter estimation for Lucas-Kanade optical flow Sensors 12 (9), 12694–12709.

Shi, X., Huang, Z., Bian, W., Li, D., Zhang, M., Cheung, K.C., See, S., Qin, H., Dai, J., Li, H., 2023. Videoflow: Exploiting temporal cues for multi-frame optical flow estimation In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12469–12480.

Shi, H., Zhou, Y., Yang, K., Ye, Y., Yin, X., Yin, Z., Meng, S., Wang, K., 2022. PanoFlow: Learning optical flow for panoramic images arXiv preprint arXiv:2202.13388.

Shugrina, M., Liang, Z., Kar, A., Li, J., Singh, A., Singh, K., Fidler, S., 2019. Creative flow+ dataset In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5384–5393.

Stahl, A., Schellewald, C., Stavdahl, O., Aamo, O.M., Adde, L., Kirkerod, H., 2012. An optical flow-based method to predict infantile cerebral palsy IEEE Trans. Neural Syst. Rehabil. Eng. 20 (4), 605–614. http://dx.doi.org/10.1109/TNSRE.2012.2195030.

Sun, D., Roth, S., Black, M.J., 2010. Secrets of optical flow estimation and their principles In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, pp. 2432–2439.

Sun, D., Roth, S., Black, M.J., 2014. A quantitative analysis of current practices in optical flow estimation and the principles behind them Int. J. Comput. Vis. 106, 115–137.

Sun, D., Vlasic, D., Herrmann, C., Jampani, V., Krainin, M., Chang, H., Zabih, R., Freeman, W.T., Liu, C., 2021. AutoFlow: Learning a better training set for optical flow In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 10093–10102.

Sun, D., Yang, X., Liu, M.-Y., Kautz, J., 2018. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 8934–8943.

Teed, Z., Deng, J., 2020. Raft: Recurrent all-pairs field transforms for optical flow In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16. Springer, pp. 402–419.

Tomasi, C., Manduchi, R., 1998. Bilateral filtering for gray and color images In: Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271). pp. 839–846. http://dx.doi.org/10.1109/ICCV.1998.710815.

Trinh, D.-H., Daul, C., 2019. On illumination-invariant variational optical flow for weakly textured scenes Comput. Vis. Image Underst. 179, 1–18.

Tu, Z., Li, H., Xie, W., Liu, Y., Zhang, S., Li, B., Yuan, J., 2022a. Optical flow for video super-resolution: A survey http://dx.doi.org/10.48550/ARXIV.2203.10462, arXiv URL https://arxiv.org/abs/2203.10462.

Tu, Z., Li, H., Xie, W., Liu, Y., Zhang, S., Li, B., Yuan, J., 2022b. Optical flow for video super-resolution: a survey Artif. Intell. Rev. 55 (8), 6505–6546.

Tu, Z., Van Der Aa, N., Van Gemeren, C., Veltkamp, R.C., 2014. A combined post-filtering method to improve accuracy of variational optical flow estimation Pattern Recognit. 47 (5), 1926–1940.

Tu, Z., Xie, W., Zhang, D., Poppe, R., Veltkamp, R.C., Li, B., Yuan, J., 2019. A survey of variational and CNN-based optical flow techniques Signal Process., Image Commun. 72, 9–24.

Van de Weijer, J., Gevers, T., Geusebroek, J.-M., 2005. Edge and corner detection by photometric quasi-invariants IEEE Trans. Pattern Anal. Mach. Intell. 27 (4), 625–630.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need Adv. Neural Inf. Process. Syst. 30.

Vedula, S., Baker, S., Rander, P., Collins, R., Kanade, T., 1999. Three-dimensional scene flow In: Proceedings of the Seventh IEEE International Conference on Computer Vision, Vol. 2. IEEE, pp. 722–729.

Wang, Y., Yang, Y., Yang, Z., Zhao, L., Wang, P., Xu, W., 2018. Occlusion aware unsupervised learning of optical flow In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4884–4893.

Weinzaepfel, P., Revaud, J., Harchaoui, Z., Schmid, C., 2013. DeepFlow: Large displacement optical flow with deep matching In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1385–1392.

Welch, G., Bishop, G., 1995. An Introduction to the Kalman Filter. University of North Carolina at Chapel Hill.

Wulff, J., Black, M.J., 2015. Efficient sparse-to-dense optical flow estimation using a learned basis and layers In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 120–130.

Xiao, J., Cheng, H., Sawhney, H., Rao, C., Isnardi, M., 2006. Bilateral filtering-based optical flow estimation with occlusion detection In: Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9. Springer, pp. 211–224.

Xiao, F., Lee, Y.J., 2016. Track and segment: An iterative unsupervised approach for video object proposals In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 933–942. http://dx.doi.org/10.1109/CVPR.2016.107.

Xu, L., Jia, J., Matsushita, Y., 2011. Motion detail preserving optical flow estimation IEEE Trans. Pattern Anal. Mach. Intell. 34 (9), 1744–1757.

Xu, L., Lu, C., Xu, Y., Jia, J., 2012. Structure extraction from texture via relative total variation ACM Trans. Graph. 31 (6).

Xu, C., Pilla, J.J., Isaac, G., Gorman, J.H., Blom, A.S., Gorman, R.C., Ling, Z., Dougherty, L., 2010. Deformation analysis of 3D tagged cardiac images using an optical flow method J. Cardiovasc. Magn. Reson. 12, 1–14.

Xu, J., Ranftl, R., Koltun, V., 2017. Accurate optical flow via direct cost volume processing In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1289–1297.

Yang, Z., Simon, R., Li, Y., Linte, C.A., 2021. Dense depth estimation from stereo endoscopy videos using unsupervised optical flow methods In: Medical Image Understanding and Analysis: 25th Annual Conference, MIUA 2021, Oxford, United Kingdom, July 12–14, 2021, Proceedings 25. Springer, pp. 337–349.

Yin, X.L., Liang, D.X., Wang, L., Xu, J., Han, D., Li, K., Yang, Z.Y., Xing, J.H., Dong, J.Z., Ma, Z.Y., 2022. Optical flow estimation of coronary angiography sequences based on semi-supervised learning Comput. Biol. Med. 146, 105663.

Yin, X.C., Zuo, Z.Y., Tian, S., Liu, C.L., 2016. Text detection, tracking and recognition in video: A comprehensive survey IEEE Trans. Image Process. 25 (6), 2752–2773. http://dx.doi.org/10.1109/TIP.2016.2554321.

Yu, J.J., Harley, A.W., Derpanis, K.G., 2016. Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness In: Computer Vision–ECCV 2016 Workshops: Amsterdam, the Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14. Springer, pp. 3–10.

Zabih, R., Woodfill, J., 1994. Non-parametric local transforms for computing visual correspondence In: European conference on computer vision. Springer, pp. 150–158.

Zach, C., Pock, T., Bischof, H., 2007. A duality based approach for realtime tv-l 1 optical flow In: Pattern Recognition: 29th DAGM Symposium, Heidelberg, Germany, September 12-14, 2007. Proceedings 29. Springer, pp. 214–223.

Zachiu, C., Papadakis, N., Ries, M., Moonen, C., de Senneville, B.D., 2015. An improved optical flow tracking technique for real-time MR-guided beam therapies in moving organs Phys. Med. Biol. 60 (23), 9003. http://dx.doi.org/10.1088/0031-9155/60/23/9003.

Zeng, X., Zhao, X., Zhong, X., Liu, G., 2023. A survey of micro-expression recognition methods based on LBP, optical flow and deep learning Neural Process. Lett. 1–32.

Zhai, M., Xiang, X., Lv, N., Kong, X., 2021. Optical flow and scene flow estimation: A survey Pattern Recognit. 114, 107861.

Zhang, C.-x., Chen, Z., Li, M., 2016. Review of the 3D reconstruction technology based on optical flow of monocular image sequence Acta Electon. Sinica 44 (12), 3044.

Zhao, S., Sheng, Y., Dong, Y., Chang, E.I., Xu, Y., et al., 2020. Maskflownet: Asymmetric feature matching with learnable occlusion mask In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6278–6287.

Zheng, Y., Zhang, M., Lu, F., 2020. Optical flow in the dark In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 6748–6756. http://dx.doi.org/10.1109/CVPR42600.2020.00678.

Zickler, T., Mallick, S.P., Kriegman, D.J., Belhumeur, P.N., 2008. Color subspaces as photometric invariants Int. J. Comput. Vis. 79, 13–30.