Real-World Evaluation of Optical Flow with Varying Lighting Conditions

Zhuoyue Ge[1]

Supervisor(s): Jan van Gemert[1]**, Sander Gielisse**[1]

[1]**EEMCS, Delft University of Technology, The Netherlands**

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 22, 2025

Name of the student: Zhuoyue Ge
Final project course: CSE3000 Research Project
Thesis committee: Jan van Gemert, Sander Gielisse, Alexios Voulimeneas

Figure 1: Optical flow visualization showing two frames of a sequence, ground-truth optical flow (color coded), and the color code to read the vector at each pixel.[1]

Abstract

Optical flow estimation is a core task in computer vision, yet many existing models struggle with lighting-induced appearance changes that are common in real-world scenarios. This work presents a focused evaluation of recent deep learning-based optical flow models under controlled lighting variations, using a custom dataset composed of indoor and outdoor scenes recorded with a static camera. Scenarios include glare, moving shadows, intensity shifts, and outdoor shadows, with ground truth flow defined as zero to isolate the effect of illumination changes. Four models—RAFT, GMFlow, SEA-RAFT, and FlowDiffuser—are benchmarked using standard metrics (EPE and F1-all). The results reveal that even in the absence of physical motion, several models produce significant flow estimates, particularly under shadow and intensity variation. SEA-RAFT and RAFT show relatively higher robustness, while GMFlow and FlowDiffuser are more sensitive to lighting artifacts. The findings highlight a critical gap in current model generalization and emphasize the need for lighting-aware architectures and training strategies.

1 Introduction

Optical flow, defined as the apparent motion of brightness patterns in an image sequence [1], is a fundamental concept in computer vision. It captures the pixel-level movement of objects between consecutive frames in a video, enabling the analysis of dynamic scenes. An example of optical flow visualization is shown in Fig. 1, where the third panel illustrates the standard color-coded representation: hue encodes the direction of each flow vector, and intensity indicates its magnitude.

Accurate estimation of optical flow is important for various applications, including autonomous driving [2] and video analysis [3]. Recent deep learning models such as RAFT [4], GMFlow [5], SEA-RAFT [6], and FlowDiffuser [7] have shown strong results, but they are mainly trained and evaluated on synthetic datasets, where ground truth flow is available through rendering.

While synthetic datasets allow for large-scale training, they do not fully capture the complexity of real-world scenes. As a result, there is a significant "sim-to-real gap" – models perform well on synthetic benchmarks but often struggle in real-world scenarios [8; 9]. Although several works have introduced more realistic benchmarks [2; 9; 8], systematic evaluation across diverse real-world conditions is still limited.

Lighting variation is one of the most challenging and underexplored aspects of the sim-to-real gap. Real-world environments frequently experience unpredictable changes in illumination, such as intensity shifts, glare, and dynamic shadows [10]. Accurate estimation under lighting variation is especially critical in safety-sensitive applications like autonomous driving [2], surveillance [11], and robotics [12]. For instance, when a vehicle drives through a tunnel or under trees on a sunny day, rapid shifts in lighting can create shadows or flickering patterns on the road surface. These artifacts can be misinterpreted as motion, leading to incorrect optical flow predictions and compromising downstream tasks like object detection and navigation [2]. Therefore, robust handling of lighting variation is vital for reliable real-world performance. However, these lighting complexities are rarely reflected in synthetic datasets, which typically rely on fixed lighting setups and struggle to reproduce phenomena like sudden glare, limiting their realism and generalizability [10].

This sim-to-real gap under challenging lighting conditions raises the central research question of this paper: How well do optical flow models evaluated on synthetic datasets perform in real-world scenarios with varying lighting conditions? Specifically, which lighting conditions are most likely to cause model failures, and how do different models respond to these conditions?

To address these questions, this paper introduces a new dataset based on real-world video, specifically designed to include a wide range of lighting conditions. Utilizing this dataset, a systematic evaluation of state-of-the-art optical flow models is conducted to assess their performance and robustness under different lighting variations. The study further analyzes the failure modes of these models when exposed to diverse lighting conditions, providing a comparative overview of their strengths and weaknesses.

The remainder of this paper is organized as follows: Section 2 reviews related work; Section 3 details the methodology and main contribution; Section 4 presents the experimental setup and results ; Section 5 discusses responsible research considerations; Section 6 offers a broader discussion of the results; and Section 7 concludes the paper and outlines directions for future research.

2 Background

Optical Flow models Optical flow estimation is a long-standing problem in computer vision, aiming to capture pixel-level motion between consecutive frames. The field has evolved significantly – from classical variational methods to modern deep learning-based models.

Early approaches such as the Horn-Schunck algorithm [1] and the Lucas-Kanade method [13] rely on assumptions like brightness constancy and spatial smoothness. While these methods laid the foundation for optical flow, they often fail in complex scenes involving occlusions, large displacements, or significant lighting variation.

In recent years, deep learning has brought remarkable improvements in both accuracy and robustness. Pioneering models like FlowNet [14] and PWC-Net [15] introduced convolutional neural networks to estimate optical flow directly from image pairs. Building on these ideas, more advanced architectures such as RAFT [4], GMFlow [5], SEA-RAFT [6], and FlowDiffuser [7] have achieved state-of-the-art performance through iterative refinement, global matching, and attention mechanisms.

Most state-of-the-art optical flow models are trained and evaluated primarily on synthetic datasets such as FlyingChairs [14] and FlyingThings3D [16]. These datasets are generated through computer graphics rendering pipelines, where objects and scenes are artificially created, and pixel-level motion between frames can be computed automatically from the rendering engine. This enables the generation of large-scale datasets with accurate ground truth optical flow, making them ideal for supervised learning. However, synthetic datasets inherently fall short in representing the complexities of real-world scenarios. They often fail to capture challenging phenomena like non-rigid motion (e.g., human movement) and dynamic lighting variations, which lead to the sim-to-real gap.

To overcome this gap, several real-world datasets have been developed, including KITTI [2] and MPI Sintel [17]. Being one of the most widely used, KITTI features street scenes captured from a moving vehicle, providing a more realistic testbed for evaluating optical flow in autonomous driving contexts. Ground truth in KITTI is generated using a combination of LIDAR measurements and manual annotations. While this approach brings the evaluation closer to real-world conditions, it still presents limitations. The ground truth is sparse and may introduce noise or inaccuracies [4]. Moreover, KITTI contains a limited range of lighting conditions, as most data was collected under relatively consistent outdoor settings.

To address these limitations, this paper introduces a new real-world optical flow dataset with dense annotation that explicitly focuses on diverse and challenging lighting conditions. Using this dataset, a comprehensive evaluation of several recent deep learning-based optical flow models – RAFT [4], GMFlow [5], SEA-RAFT [6], and FlowDiffuser [7] – is conducted to investigate their robustness under realistic lighting changes.

Varying Lighting Conditions Lighting variation remains a major challenge in optical flow estimation [18]. Changes in illumination can significantly affect pixel intensity values, which many optical flow algorithms rely on for matching corresponding points across frames. As a result, even minor lighting fluctuations can lead to large estimation errors, particularly in real-world, uncontrolled environments.

Traditional optical flow models have made early attempts to mitigate the impact of illumination changes. For instance, methods based on brightness constancy often fail under varying lighting, prompting researchers to incorporate more robust features. Tsenov et al. [18] proposed image preprocessing techniques, such as histogram equalization and gradient-based normalization, to improve performance in dynamic lighting conditions for real-time applications. Other classical methods leverage gradient orientation or texture-based cues, which are less sensitive to lighting changes than raw intensity values [19]. However, these approaches typically trade off motion sensitivity for lighting invariance, which can limit their applicability in complex environments.

Modern deep learning-based optical flow models have brought significant improvements in robustness and accuracy, often credited to large-scale supervised training and heavy data augmentation. Many models, such as FlowNet [14], RAFT [4], and GMFlow [5], employ photometric augmentations like brightness jittering, contrast shifts, and color noise to simulate lighting variability. While these techniques improve generalization to some extent, they still fall short of capturing the full range of illumination dynamics encountered in real-world environments.

Inspired by prior work on handling lighting variations for tradition models [18], this paper uses a static-camera setup while intentionally introducing dynamic lighting conditions, for example, toggling a light source to create sudden intensity variations. Several realistic scenarios – such as dynamic shadows, glare, and lighting intensity changes – are included to systematically examine how each model responds under stress. This approach enables deeper analysis of failure modes and highlights areas where current models still struggle.

3 Methodology

To evaluate the robustness of state-of-the-art optical flow models under varying lighting conditions, this work introduces a practical evaluation pipeline focused on real-world data collection, frame selection, model benchmarking, and detailed performance analysis.

[1] Image from Torralba, Isola, and Freeman, *Foundations of Computer Vision*, 2024. Used under CC BY-NC-ND license. Source: https://visionbook.mit.edu/optical_flow.html.

3.1 Dataset Collection

Video sequences are recorded in real-world environments using a stationary camera setup. Both indoor and outdoor scenes are included to capture a broad spectrum of natural lighting variations. The scenes are intentionally selected to incorporate challenging lighting conditions, such as dynamic shadows, changing lights intensity, and glares. These conditions are known to degrade optical flow accuracy and are chosen to reflect the complexities encountered in everyday visual tasks [20].

3.2 Frame Selection

To extract representative data from the recorded videos, a custom frame selection tool is developed. This tool provides an interactive user interface that allows manual selection of frame pairs deemed most illustrative of specific lighting changes. The selected pairs are then exported in a KITTI-compatible format, making it straightforward to evaluate with standard optical flow metrics and existing benchmarking tools.

3.3 Model Evaluation

Four state-of-the-art optical flow models – RAFT [4], MFlow [5], SEA-RAFT [6], and FlowDiffuser [7] – are benchmarked on the collected dataset. These models span four key design paradigms—iterative refinement, global matching, adaptive enhancement, and generative diffusion—making them an ideal set for evaluating how architectural choices affect robustness to lighting-induced challenges. Each model is tested on the collected dataset, with performance assessed across various lighting scenarios. This evaluation allowed for a comprehensive comparison of model architecture on robustness to lighting-induced artifacts, and provides a comparative analysis of failure cases.

3.4 Evaluation Metrics

To quantitatively assess model performance, two standard optical flow evaluation metrics are employed: End-Point Error (EPE) and F1 score. The End-Point Error (EPE) computes the average Euclidean distance between the predicted optical flow vectors and the ground truth vectors across all pixels [21]. It provides a direct measure of overall accuracy in estimating motion. The F1 score, often referred to as F1-all, measures the percentage of pixels whose EPE exceeds a predefined threshold – typically 3 pixels or 5% of the magnitude of the ground truth flow vector, whichever is larger [2]. This metric reflects the proportion of significantly incorrect flow predictions, offering insight into the reliability and robustness of each model, especially under challenging lighting conditions.

3.5 Analysis of Failure Modes

Alongside quantitative evaluation, a qualitative analysis is carried out to identify recurring failure modes in the optical flow predictions. This involved a careful visual examination of model outputs under diverse lighting conditions to reveal specific challenges—such as moving shadows, glare, and inconsistent illumination – that tend to degrade performance. Identifying these patterns is critical for informing the design of more robust and lighting-aware optical flow algorithms.

4 Experiments

4.1 Data Collection

To evaluate the effect of lighting changes on optical flow estimation, a custom dataset is recorded using the iPhone 12's default Wide Camera (26mm, f/1.6) at a resolution of 1440×1920 (1440p) and frame rate of 30 fps. The device is mounted vertically on a stationary tripod with a clamp to eliminate camera shake or movement throughout the recording. Automatic exposure, focus, and white balance settings are left enabled to reflect standard consumer recording conditions.

For indoor scenes, recordings are conducted at night to fully control the lighting environment and eliminate natural light interference. A corner of a residential living room is selected, containing a consistent background of small bookshelves, one floor lamp, one table lamp, and decorative items. Artificial lighting is used exclusively, and no objects are moved between sessions to ensure that the only variable is the lighting condition. Three distinct lighting scenarios are designed: glare, moving shadows, and lighting intensity changes.

Glare To simulate glare artifacts, a bright light source (iPhone flashlight from a second phone) is used to introduce direct light near the camera lens. The flashlight is gradually moved toward the edge of the camera's field of view until lens flare artifacts (glare rings) appeared. The flashlight itself should not appear in the scene. Other room lights remained constant throughout to isolate the glare effect. An example is shown in the first pair of frames in Figure 2.

Moving Shadows The moving shadows scenario is designed to create dynamic shading effects without any physical object motion. The iPhone flashlight is again used as a single-point light source. Two methods are used to introduce motion into the shadows: (1) Occlusion Movement - placing a moving object between the light source and the background to cast dynamic shadows; (2)Light Source Angle Change - changing the flashlight angle to alter the position of static shadows. Light sources within the scene remained stationary in both cases. This setup is illustrated in the second frame pair of Figure 2.

Lighting Intensity Changes Multiple lighting sources, including a floor lamp, a table lamp, and an adjustable ceiling light, are used to simulate changes in ambient lighting. The ceiling light featured a built-in dimmer, allowing for smooth transitions in brightness. Lighting variations are introduced in two ways: (1) toggling individual lamps on and off to create discrete changes in illumination, and (2) gradually adjusting the brightness of the dimmable ceiling light to simulate continuous lighting intensity changes. Throughout these recordings, the spatial placement of all lamps and scene objects re-

Figure 2: Paired frames illustrating four lighting-related visual phenomena: (1) Glare, (2) Moving Shadows, (3) Lighting Intensity Changes, and (4) Outdoor Shadows. Each pair consists of two frames captured under different lighting conditions, displayed at their original scale.[2]

mained fixed to ensure that only lighting intensity varied between frames. The third frame pair in Figure 2 demonstrates this lighting effect.

For all indoor settings, only the lighting environment is manipulated; no objects in the scene are moved.

Outdoor Shadows Outdoor recordings are captured on sunny days across ten different locations. These scenes included natural lighting variations, such as moving shadows cast by tree leaves on grass and pavement. The fourth frame pair in Figure 2 depicts a typical outdoor shadow case.

4.2 Data Adjustment and Annotation

To ensure consistency with established benchmarks and facilitate fair model evaluation, all video sequences are first rotated to a horizontal orientation and downsampled to a resolution of 1242×375 pixels – the standard resolution used in the KITTI dataset [9]. This resolution preserves important spatial details while ensuring compatibility with widely used optical flow evaluation frameworks.

For the indoor dataset, a total of 20 frame pairs are manually selected for each targeted lighting condition. Selection is performed using the custom-built frame selection tool, with careful attention to ensuring that each pair distinctly exhibits the intended lighting effect. This targeted selection enables fine-grained analysis of model behavior under specific visual stressors.

For the outdoor dataset, an automatic frame selection strategy is adopted to better reflect real-world deployment scenarios. From each of the ten locations, the first 20 frame pairs are selected at fixed intervals of 10 frames, such as frames 0 and 10, 10 and 20, 20 and 30, and so on. This sampling strategy is designed to highlight gradual lighting changes, particularly those caused by moving shadows, while ensuring the physical scene remains static with no object or camera motion.

Across all scenes, the ground truth optical flow is defined as zero, as no actual displacement occurs between frames. This setting is intentionally designed to test whether models incorrectly predict motion in the presence of dynamic lighting.

4.3 Models Used

Four state-of-the-art optical flow models are evaluated using their publicly available PyTorch implementations and pre-trained weights:

- RAFT [4]: Evaluated using the pre-trained model `raft-things.pth`, trained on the FlyingThings3D dataset [16].
- GMFlow [5]: Evaluated using the pre-trained model `gmflow_kitti-285701a8.pth`, trained with the KITTI dataset [9].
- SEA-RAFT [6]: Evaluated using the pre-trained model `Tartan-C-T-TSKH-kitti432x960-M.pth`, trained with KITTI-style data.
- Flow Diffuser [7]: Evaluated using the pre-trained model `FlowDiffuser-things.pth`, trained with the FlyingThings3D dataset [16].

All models are tested using the default evaluation script provided by the authors, without any fine-tuning or post-processing modifications.

4.4 Results

Table 1 presents the average End-Point Error (EPE) and F1-all scores for each model under four different lighting scenarios: glare, moving shadows, lighting intensity changes (indoor), and outdoor shadows. These results represent the mean performance across all frame pairs per condition.

Overall, SEA-RAFT [6] and RAFT [4] demonstrate the most consistent robustness to lighting-induced changes, achieving the lowest average EPE and F1-all scores across most conditions. In contrast, FlowDiffuser [7] and GMFlow [5] exhibit significantly higher error rates in scenarios involving dynamic shadows and intensity changes, indicating greater sensitivity to non-motion-based appearance changes.

Under glare, all models maintain relatively low error, with SEA-RAFT achieving the best performance (EPE: 0.26, F1-all: 0.58). RAFT and FlowDiffuser also perform comparably well, while GMFlow shows notably higher error, suggesting reduced resilience to high-contrast highlights.

The moving shadows condition poses substantial difficulty. FlowDiffuser records the highest error (EPE: 28.56, F1-all: 42.77), followed by GMFlow. These results indicate that both models tend to misinterpret illumination changes as motion. SEA-RAFT and RAFT perform significantly better, with SEA-RAFT achieving the lowest average EPE (4.09), highlighting its improved handling of shadow-induced variation.

Model	Glare		Moving Shadows		Lighting Intensity		Outdoor Shadows	
	EPE	F1-all	EPE	F1-all	EPE	F1-all	EPE	F1-all
RAFT	0.31	0.13	13.35	14.43	12.24	16.35	1.94	3.01
GMFlow	3.03	5.36	7.15	34.90	4.35	28.67	2.57	7.21
SEA-RAFT	0.26	0.58	4.09	14.94	6.84	13.16	1.53	3.34
FlowDiffuser	0.37	0.39	28.56	42.77	22.89	30.92	3.92	6.16

Table 1: Comparison of EPE and F1-all scores in indoor and outdoor lighting conditions.

In the lighting intensity change scenario, RAFT yields the most stable results (EPE: 12.24, F1-all: 16.35), while GMFlow and FlowDiffuser experience a marked increase in error. SEA-RAFT performs moderately well, though still impacted by sudden intensity shifts.

Under outdoor shadows, all models experience a slight decline in performance. SEA-RAFT again achieves the lowest average EPE (1.53), while FlowDiffuser and GMFlow show increased sensitivity to natural shadow movement, reflected in their higher F1-all scores.

It is worth noting that although the reported values represent per-condition averages, the within-condition variance is often substantial. Certain frame pairs—particularly those involving rapid lighting transitions or complex shadow patterns—lead to significantly higher error rates. This variability underscores the importance of complementary qualitative analysis, which is addressed in the following section.

5 Responsible Research Considerations

5.1 Ethical Risks

This work investigates optical flow models under lighting changes using static scenes without any human subjects, animals, or identifiable personal data. As such, it poses minimal ethical risk. No deceptive methods or manipulations were involved, and all data was collected in controlled indoor environments or public outdoor locations where privacy concerns were carefully considered. The primary research goal is technical: to assess model sensitivity to appearance-based scene changes, not to make inferences about human behavior or identity.

5.2 Dataset Privacy

All data was collected by the authors using a personal smartphone. Indoor scenes were recorded in a private residence with no personally identifiable items visible, and no individuals appeared in any footage. Outdoor scenes were captured in public spaces (e.g., parks, sidewalks), ensuring that no people, license plates, or private property were included in the recordings. The dataset contains only non-sensitive visual content focused on static scenes and natural or artificial lighting effects.

To further ensure privacy, all video frames were reviewed prior to annotation and publication. Any footage inadvertently containing private information was discarded.

5.3 Reproducibility

This study was designed with reproducibility in mind. All details of the recording setup, including hardware (iPhone 12 Wide Camera), resolution (1440×1920 down-sampled to 1242×375), lighting arrangements, and frame sampling strategies, are fully documented in Section 4. The annotation procedure is described in detail, along with the choice of zero ground truth flow based on the static nature of the scenes.

The following materials will be made publicly available upon acceptance of this work:

- The full dataset (all extracted frame pairs with frame annotations in KITTI-compatible format)
- Source code for frame selection tools used in the analysis

These materials will be released under a permissive license for research use and hosted on an open repository to support reproducibility and further research. The dataset and tools will be available at: https://github.com/aufos/Optical-flow-frame-section-tool.git

5.4 Bias and Generalization

Although the dataset is diverse in lighting conditions, it is geographically and environmentally limited to a single indoor setting and ten outdoor scenes in one region. As such, it may not capture the full range of lighting behaviors in different environments, seasons, or cultures. Future extensions may involve broader geographic sampling or synthetic augmentation to increase diversity.

This study also highlights an important bias in current optical flow models – their tendency to interpret lighting changes as physical motion. Identifying and mitigating this bias is a key motivation behind the work and can help guide the development of more robust, perception-aware models.

5.5 Environmental Impact

This work used pre-trained models without fine-tuning and operated on relatively small-resolution videos. Inference and evaluation were performed on a single GPU workstation. No large-scale model training or hyperparameter sweeps were conducted, minimizing compute usage and energy consumption.

6 Discussion

The results presented in Table 1 highlight the varying degrees of robustness exhibited by state-of-the-art optical flow models when exposed to realistic lighting changes. While all models were evaluated in scenes without any physical motion, most still produced non-zero flow estimates in response to lighting-induced appearance changes—demonstrating a critical limitation in existing approaches.

6.1 Lighting Effects on Model Behavior

Across all tested scenarios, models demonstrated the greatest stability under glare, with relatively low error values. This suggests that glare, while visually disruptive, may not introduce large structural changes in the image that confuse modern flow estimation pipelines.

In contrast, moving shadows and lighting intensity variations caused significant performance degradation in multiple models, particularly GMFlow [5] and FlowDiffuser [7]. These models, both of which rely heavily on global matching and visual appearance cues, were more prone to misinterpreting illumination changes as motion. This indicates a lack of robustness to non-geometric transformations and highlights a common limitation among methods trained primarily on synthetic or static lighting datasets.

SEA-RAFT [6] and RAFT [4], which adopt iterative refinement and feature-level warping mechanisms, consistently performed better under varying lighting. Notably, SEA-RAFT's training on diverse and motion-consistent datasets such as TartanAir appears to contribute to its ability to suppress false motion in lighting-dominated scenes.

6.2 Limitations of Current Pretraining Strategies

The analysis also reveals the influence of training data on model generalization. Models trained on a single dataset (e.g., RAFT and FlowDiffuser on FlyingThings3D [16] and GMFlow on KITTI [9]) tend to perform worse under lighting conditions not well represented in their training distribution. SEA-RAFT, trained on a combination of real-world and synthetic datasets including TartanAir [22] and KITTI, demonstrates greater robustness—emphasizing the importance of diverse, motion-consistent training environments for improved generalization.

6.3 Need for Lighting-Aware Architectures

A key insight from this evaluation is that current architectures lack explicit mechanisms to differentiate between appearance changes due to motion and those due to lighting. Even in a static scene, changes in pixel intensity caused by shadows or lighting transitions often result in false optical flow predictions. This suggests that future models could benefit from integrating additional cues, such as temporal consistency checks, photometric invariants, or illumination-robust feature representations.

6.4 Implications for Real-World Applications

These findings have important implications for real-world applications of optical flow, such as robotics, surveillance, and autonomous driving. In such domains, lighting variations are frequent and unpredictable. Without robustness to these changes, optical flow systems risk producing misleading motion estimates, which can undermine downstream tasks like object tracking, scene understanding, and navigation.

7 Conclusions and Future Work

7.1 Conclusion

This work presents a focused evaluation of state-of-the-art optical flow models under realistic lighting variations, using a custom-constructed dataset with controlled glare, moving shadows, lighting intensity changes, and outdoor shadow effects. Despite the absence of physical motion in the test sequences, several models—particularly GMFlow [5] and FlowDiffuser [7]—produced significant flow predictions, revealing a critical vulnerability to lighting-induced appearance changes.

Among the evaluated models, SEA-RAFT [6] and RAFT [4] demonstrated comparatively strong robustness, benefiting from architectural features like iterative refinement and diverse training data. However, even these models exhibited performance degradation in more complex lighting scenarios, such as moving shadows or abrupt intensity transitions. These findings underscore the limitations of current model architectures and training strategies in handling real-world visual variability.

7.2 Future Work

Based on the findings of this study, several directions are recommended for future research.

Controlled and Dynamic Scene Simulation While this study focuses on completely static scenes, future datasets could incorporate small, controlled camera movements—such as handheld jitter or slow panning—combined with lighting changes. This would more accurately reflect real-world scenarios, where lighting variation and minor viewpoint changes often co-occur. However, such setups would require non-zero ground truth optical flow, which introduces additional complexity in annotation and calibration.

Dataset Expansion Future work should aim to construct datasets that capture the same scene under systematically varied lighting conditions, such as different times of day (morning, afternoon, evening), weather conditions (sunny, overcast), and artificial lighting setups (diffuse vs. directional). Recording multiple lighting conditions for identical scenes would allow more precise isolation of the impact of illumination on model predictions. With a larger and more diverse dataset, models could not only be evaluated but also trained on lighting-controlled data to enhance their real-world generalization.

Model Adaptation Future models could incorporate architectural components or training strategies specifically aimed at handling lighting variation. For instance, introducing modules that disentangle illumination changes from scene motion, or training with lighting-invariant features, may help reduce false positives in static scenes. Fine-tuning existing models on datasets that emphasize lighting variation may also lead to more stable performance in deployment scenarios with dynamic illumination.

By addressing these aspects, future research can build more robust and perceptually grounded optical flow models capable of handling the complexities of real-world visual environments.

References

[1] Berthold K P Horn and Brian G Schunck. Determining optical flow. *Artificial intelligence*, 17(1-3):185–203, 1981.

[2] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[3] Wenzhe Zhu, Cuiling Lan, Junliang Xing, Wenjun Zeng, and Jiaying Li. Deep learning for video classification: A review. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(10):2764–2777, 2018.

[4] Zachary Teed and Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow. In *European Conference on Computer Vision (ECCV)*, pages 402–419. Springer, 2020.

[5] Haofei Xu, Jing Zhang, Jianfei Cai, Hamid Rezatofighi, and Dacheng Tao. Gmflow: Learning optical flow via global matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8121–8130, 2022.

[6] Yihan Wang, Lahav Lipson, and Jia Deng. Sea-raft: Simple, efficient, accurate raft for optical flow. In *Computer Vision – ECCV 2024*, pages 36–54. Springer, 2025.

[7] Ao Luo, Xin Li, Fan Yang, Jiangyu Liu, Haoqiang Fan, and Shuaicheng Liu. Flowdiffuser: Advancing optical flow estimation with diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19167–19176, 2024.

[8] Daniel J. Butler, Jonas Wulff, Garrett B. Stanley, and Michael J. Black. A naturalistic open source movie for optical flow evaluation. In *European Conference on Computer Vision (ECCV)*, pages 611–625, 2012.

[9] Moritz Menze and Andreas Geiger. Object scene flow for autonomous vehicles. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3061–3070, 2015.

[10] Joel Janai, Fatma Güney, Aseem Behl, and Andreas Geiger. Computer vision for autonomous vehicles: Problems, datasets and state-of-the-art. *Foundations and Trends in Computer Graphics and Vision*, 12(1–3):1–308, 2020.

[11] P. D. Garje, M. S. Nagmode, and K. C. Davakhar. Optical flow based violence detection in video surveillance. In *2018 International Conference on Advances in Communication and Computing Technology (ICACCT)*, pages 208–212, 2018.

[12] Naoya Ohnishi and Atsushi Imiya. Visual navigation of mobile robot using optical flow and visual potential field. In Gerald Sommer and Reinhard Klette, editors, *Robot Vision*, pages 412–426, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[13] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 674–679, 1981.

[14] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick van der Smagt, Daniel Cremers, and Thomas Brox. Flownet: Learning optical flow with convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2758–2766, 2015.

[15] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8934–8943, 2018.

[16] Nikolaus Mayer, Eddy Ilg, Philip Häusser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. 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 (CVPR)*, pages 4040–4048, 2016.

[17] David J Butler, Jonas Wulff, Garrett B Stanley, and Michael J Black. A naturalistic open source movie for optical flow evaluation. In *European Conference on Computer Vision (ECCV)*, pages 611–625. Springer, 2012.

[18] Georgi Tsenov, Stanislav Obdrzalek, Andrej Ferko, Radek Smid, and Michal Sobotka. Making optic flow robust to dynamic lighting conditions for real-time operation. *Pattern Recognition Letters*, 94:228–234, 2017.

[19] Thomas Brox, Andres Bruhn, Nils Papenberg, and Joachim Weickert. High accuracy optical flow estimation based on a theory for warping. In *European conference on computer vision*, pages 25–36. Springer, 2004.

[20] Simon Baker, Daniel Scharstein, JP Lewis, Stefan Roth, Michael J Black, and Richard Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1):1–31, 2011.

[21] Jiaqi Zhao. *Learning the Loss in Optical Flow Estimation based on the End-Point Error*. PhD thesis, University of Stuttgart, 2021.

[22] Wenshan Wang, Yuxiang Yang, Xinyu Wang, Andrew Markham, and Niki Trigoni. Tartanair: A dataset to push the limits of visual slam. *arXiv preprint arXiv:2010.09733*, 2020.