Triangular Consistency as a Universal Constraint for Learning Optical Flow

Yi Xiao[1], Carlos Rodriguez Coronel[1], Jing Zhan[1], Haniyeh Ehsani Oskouie[2], Alex Wong[3], and Dong Lao[1]
Correspondence to: {yi.xiao,dong.lao}@lsu.edu

[1] Louisiana State University, Baton Rouge, LA 70803, USA
[2] University of California, Los Angeles, Los Angeles, CA 90025, USA
[3] Yale University, New Haven, CT 06520, USA

Abstract. We propose triangular consistency as a first-principled constraint for optical flow, which is agnostic to network architecture, supervision type, and dataset, and applies to both image-pair and multi-frame settings. This simple but powerful constraint is to compose two flows to induce a third flow and enforce consistency among the three. The composed flows may arise from (i) image pairs, yielding cycle consistency; (ii) multiple video frames, producing longer-range motion through temporal chaining; or (iii) image pairs combined with controlled synthetic transformations, which becomes data augmentation. This triangular consistency introduces negligible computational overhead and requires no additional annotations. Since it is derived directly from the geometry of optical flow, it does not rely on model-specific assumptions and serves as a "universal" plug-and-play component for optical flow training. Experiments show consistent improvement across supervised, unsupervised, and transfer learning settings. Code: https://github.com/lsuvision/tri-flow.

Keywords: Optical Flow · Self-Supervision · Data Augmentation

1 Introduction

The concept of optical flow was introduced by J. J. Gibson in the 1940s to describe visual motion projected onto the retina [15], decades before the emergence of computer vision. When the term was later adopted by computer scientists [17][4], our computational system inevitably converted this instantaneous quantity into a problem of inference from *discrete* frames, usually frame pairs. At first glance, this appears to be a straightforward computational discretization. However, if we consider the underlying physical process, optical flow, at its core, measures a continuous non-rigid transformation of coordinate systems of the underlying scene. Different temporal samplings of the same process must therefore obey a fundamental physical rule: they agree under composition [30, 33, 42]. If

[4] Lucas-Kanade [38] appeared slightly earlier; however, Horn-Schunck [17] is typically credited with introducing the term *optical flow* into the computer vision literature.

one mapping transports coordinates from an initial configuration to an intermediate one, and another transports them from the intermediate configuration to a final one, then their composition uniquely determines the transformation between the initial and final coordinates. Such transformations can be chained indefinitely, yet consistency always holds. As such, compositional consistency is *first-principled*: it is neither a computational assumption nor a regularization heuristic; it is directly grounded in the physics of the scene.

This knowledge by itself is nothing new. Multi-frame optical flow [13, 47, 48] and point tracking [7, 27, 33] have long used chained optical flow to improve long-range estimation. However, this compositional nature is not made explicit as a regularizing constraint. Modern learning-based optical flow methods [11, 54, 56], in contrast, are predominantly formulated as predicting displacement between two frames. While some multi-frame paradigms [10, 21, 22, 34, 49] also exploit consistency among more than two frames, they mostly build upon linear motion (constant speed) assumptions. The constraint that optical flows should agree under composition remains largely overlooked.

In this paper, we take the first step to fill this gap by studying the simplest form of composition: we consider only three frames (i.e., *triangular*): given two consecutive optical flows, their composition should match the flow directly estimated between the first and the last frames. Cycle consistency [45], where forward and backward flows between image pairs result in the identity mapping, becomes a special case, while temporal chaining of three consecutive frames and consistency under asymmetric transformations arise from the same principle. This first-principled relation provides a *universal* supervision signal: when ground truth is unavailable, it acts as a constraint for self-supervision; when ground truth is available, it enables controlled data augmentation by generating new pseudo-labels. To our knowledge, triangular consistency has not been systematically adopted by optical flow literature, and we present the first comprehensive evaluation of this principle across supervised, unsupervised, and transfer learning, spanning multiple architectures and datasets. Through the experiments, we examine how far compositional consistency can improve optical flow accuracy. Specifically, our contributions are:

- **A universal geometric constraint.** We formalize *triangular consistency*, the minimal compositional relation among three flows, and instantiate it as a training loss agnostic to architectures, supervision types, and datasets.
- **Three training regimes, one principle.** We show how the same compositional rule yields (i) temporal compositional supervision on frame triplets, (ii) cycle consistency between image pairs, and (iii) a data augmentation scheme generating pseudo-ground-truth optical flow analytically.
- **Practical integration.** The resulting training losses are lightweight, require no additional labels, and can be added to existing pipelines without modifying the estimator.
- **Consistent empirical gains.** Across transfer, unsupervised, and supervised training, triangular consistency improves accuracy and/or cross-dataset generalization, with strong gains in transfer and out-of-domain evaluation.

We observe up to 18.1% gain under single-epoch adaptation, 6-8% improvements under unsupervised training, and up to 23.1% cross-dataset gain under supervised training. In summary, triangular consistency provides a simple yet principled mechanism to improve optical flow learning. It can be readily integrated into existing training pipelines without architectural modifications or additional human supervision, yielding consistent improvements.

2 Related Work

Optical flow literature is extensive, and we only highlight some of the most relevant advancements, especially those related to consistency.

Optical flow was formulated with explicit data terms and regularization in earlier model-based and variational literature [1, 3, 5, 6, 17, 46, 52, 66–68]. Deep-learning-based methods largely inherit this framing but replace hand-crafted priors with learned matching and refinement by neural networks. Representative milestones include FlowNet [11], pyramidal warping and cost volumes in PWC-Net [54], and the strong all-pairs iterative paradigm of RAFT [56], followed by architectural variants improving either efficiency or robustness [9, 18, 51, 55, 60, 69, 74]. While these works mainly advance architectures and data scale, our focus is orthogonal: we introduce a first-principled *supervision* derived from the compositional nature of displacement fields. Therefore, it can be integrated into existing training pipelines without modifying the estimator.

Consistency as supervision signal. When ground-truth flow is unavailable, supervision is commonly derived from photometric consistency: the predicted flow warps the target image, and the discrepancy between the source image and the warped target image provides a supervision signal [45,70]. Because photometric constancy is imperfect and breaks under occlusion, later pipelines typically incorporate robust penalties, explicit occlusion reasoning, or distillation/pseudo-labeling to filter unreliable regions [21,26,35,36,40,59]. A complementary family of supervision signals is motion consistency that enforces agreement among predictions produced from different directions, frames, or transformations [23, 34, 50]. Specifically, geometric augmentation has also been explored by augmenting the source and target image with the same (*symmetric*) transformation (e.g. rotation), so that the optical flow can be augmented accordingly [34]. These methods motivate our work, but typically instantiate specific proxy constraints tied to particular training setups. In contrast, we derive a single compositional rule as a universal constraint that subsumes cycle consistency while generalizing to temporal chaining and *asymmetric* augmentation.

Composition, chaining, and long-range correspondence. Composition of displacement fields has long been exploited when moving beyond two-frame estimation [29,31,58,75]. Layered motion models provide a principled view of why composition may fail at occlusion boundaries and how multiple motion modes coexist [30,32,53], and recent benchmarks further emphasize multi-layer and non-Lambertian challenges [61]. In parallel, recent advances in dense tracking revisit optical flow as a building block for long-term trajectories through chaining and

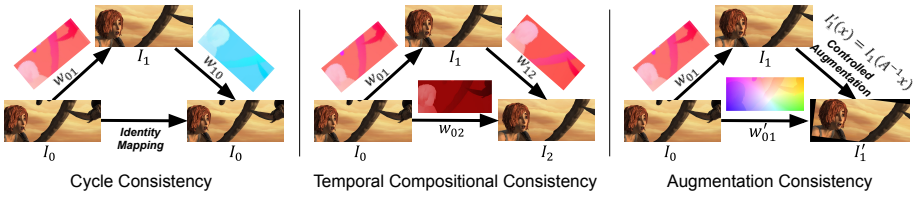

Fig. 1: Triangular consistency as a universal compositional principle. Optical flow fields compose. Triangular consistency enforces that the flow between two frames agrees with the composition of intermediate flows. Cycle consistency arises as a special case when the composition returns to the same frame (*left*). The same principle naturally generalizes to video through temporal chaining across multiple frames (*middle*), and further extends to synthetic transformations, enabling controlled data augmentation for optical flow without additional annotations (*right*).

refinement [7, 8, 27, 33, 42, 57]. These works primarily employ flow composition to initialize correspondence. Our contribution differs in scope and objective. We formulate composition as an explicit supervision signal through the simplest non-trivial compositional structure, a *triangle* (Fig. 1), and evaluate it across supervised, unsupervised, and adaptation settings. By treating composition as a universal learning constraint rather than solely as a tracking strategy, triangular consistency introduces a principled supervision mechanism that has not been explicitly incorporated into the optical flow training pipeline.

3 Method

3.1 Formalization

Let $I_t : \Omega \to \mathbb{R}^k$ ($k = 3$ for RGB) denote an image at time t, where $\Omega \subset \mathbb{R}^2$ is the image domain. Let $v_{t,t+1} : \Omega \to \mathbb{R}^2$ denote the optical flow from frame t to $t+1$, and define the corresponding warp $w_{t,t+1}(x) = x + v_{t,t+1}(x)$. Similarly, for any pair (t, s) we write $w_{t,s}(x) = x + v_{t,s}(x)$.

Compositional Structure. As in Sec. 1, we view displacement fields as non-rigid coordinate transformations and therefore compose. Given three frames I_t, I_{t+1}, I_{t+2} from the same scene, the composed warp from t to $t+2$ is

$$\widetilde{w}_{t,t+2}(x) = w_{t+1,t+2}\big(w_{t,t+1}(x)\big). \tag{1}$$

In displacement form (i.e. optical flow vectors) this becomes

$$\widetilde{v}_{t,t+2}(x) = v_{t,t+1}(x) + v_{t+1,t+2}\big(x + v_{t,t+1}(x)\big). \tag{2}$$

Triangular consistency requires that this composed flow agree with the directly estimated flow: $v_{t,t+2}(x) \approx \widetilde{v}_{t,t+2}(x)$, which expresses the simplest non-trivial compositional relation among three frames. As shown in Fig. 1, cycle consistency is a special case where the composition returns to the same frame, while temporal chaining and controlled augmentation arise from the same rule.

Triangular Consistency Loss. We thus define the triangular residual at pixel x as $r_{t,t+1,t+2}(x) = v_{t,t+2}(x) - \widetilde{v}_{t,t+2}(x)$. To mitigate the effect of outliers, we penalize it with a robust norm $\rho(\cdot)$, yielding triangular consistency loss

$$\mathcal{L}_{\text{tri}} = \sum_{x \in \Omega} M_{t,t+1,t+2}(x)\, \rho(\|r_{t,t+1,t+2}(x)\|_2)\,, \tag{3}$$

where $M_{t,t+1,t+2}(x) \in [0,1]$ is a validity mask discussed below.

Occlusion. In 2D images, pixels may become occluded or disoccluded between frames. In such regions, the mapping $w_{t,t+2}$ is not the composition of visible correspondences unless a layered motion model is available [20,30,53]. Since typical optical flow methods do not explicitly maintain a multi-layer representation, compositional consistency is violated at occlusion boundaries.

To address this, we construct $M_{t,t+1,t+2}(x)$ using forward-backward consistency checks. Regions that violate such consistency constraints are downweighted. This ensures that triangular consistency is enforced primarily on regions where valid motion correspondences can be established across frames.

3.2 Learning Objective

Cycle Consistency. When $(i,j,k) = (t, t+1, t)$, triangular consistency reduces to forward-backward cycle consistency: $v_{t,t+1}(x) + v_{t+1,t}(x + v_{t,t+1}(x)) \approx 0$. The corresponding loss is $\mathcal{L}_{\text{cyc}} = \sum_x \rho(\|v_{t,t+1}(x) + v_{t+1,t}(x + v_{t,t+1}(x))\|_2)$.

Temporal Compositional Consistency. For three distinct frames $(i,j,k) = (t, t+1, t+2)$, we enforce $v_{t,t+2}(x) = v_{t,t+1}(x) + v_{t+1,t+2}(x + v_{t,t+1}(x))$, leading to $\mathcal{L}_{\text{temp}} = \sum_x \rho(\|v_{t,t+2}(x) - \widetilde{v}_{t,t+2}(x)\|_2)$.

Augmentation Consistency. Let A be a known affine transformation. Given $I_1' = A(I_1)$, the induced analytical flow from I_1 to I_1' is $v_{1,1'}(x) = A(x) - x$. Triangular consistency enforces $v_{0,1'}(x) = v_{0,1}(x) + v_{1,1'}(x + v_{0,1}(x))$. The loss is $\mathcal{L}_{\text{aug}} = \sum_x \rho(\|v_{0,1'}(x) - v_{0,1}(x) - v_{1,1'}(x + v_{0,1}(x))\|_2)$.

Complete Objective. Let $\mathcal{L}_{\text{base}}$ denote the baseline loss (supervised loss, photometric reconstruction, flow smoothness, etc.). The overall objective becomes

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{base}} + \lambda_{\text{cyc}}\mathcal{L}_{\text{cyc}} + \lambda_{\text{temp}}\mathcal{L}_{\text{temp}} + \lambda_{\text{aug}}\mathcal{L}_{\text{aug}}. \tag{4}$$

As such, the training objective Eq. (4) is architecture-agnostic and can be directly integrated into existing optical flow models without modification, functioning as a plug-and-play supervision component.

3.3 Implementations

Flow Composition. In Eq. (2), the term $v_{t+1,t+2}\big(x + v_{t,t+1}(x)\big)$ requires evaluating $v_{t+1,t+2}$ at displaced coordinates. This is implemented by bilinear interpolation on a coordinate grid. Coordinates that fall outside image bounds are treated as invalid and excluded from the loss. Importantly, we do not warp images. Instead, the composition is applied directly to the flow field. As a result,

the supervision depends purely on the geometric relation between coordinates and is independent of color, texture, or image-feature consistency.

Controlled Augmentation. For augmentation consistency, we sample a transformation consisting of translation $t = (t_x, t_y)^\top$, rotation θ, and scale s. Given the image center $c = (c_x, c_y)^\top$, the forward transform from I_1 to I_1' is

$$A(x) = R_s(x - c) + c + t, \qquad R_s = s \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{5}$$

For resampling, we use the corresponding inverse map $A^{-1}(x) = Mx + b$, where

$$M = R_s^{-1} = \frac{1}{s} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \qquad b = c - M(c + t). \tag{6}$$

This gives an exact inverse mapping, and I_1' can be sampled from I_1 accordingly.

Now let v_{01} be the optical flow from I_0 to I_1, either from the ground truth or model prediction. Define base pixel coordinates x and transported coordinates $y = x + v_{01}(x)$. The augmented target position is

$$y' = A(y). \tag{7}$$

The induced flow from I_0 to the augmented frame I_1' is then

$$v_{01}'(x) = y' - x. \tag{8}$$

Eq. (8) is entirely analytical and computed in closed form. The augmented flow v_{01}' is thus obtained through direct coordinate transformation and does not require any image resampling or grid-based interpolation. Consequently, the supervision signal is free from interpolation artifacts regardless of the magnitude or complexity of the sampled transformation.

Importantly, although the augmented image may contain invalid regions due to pixels mapping outside the image domain, the induced optical flow field remains well-defined for *every* pixel in the source frame. The benefit of this property is non-trivial for supervising optical flow: the augmented ground-truth flow preserves all the spatial regularities and smoothness of the original displacement field without invalid pixels at image boundaries or artifacts caused by resampling. As a result, triangular augmentation remains free from artifacts even under aggressive transformations.

Occlusion Filtering. Occlusion violates cycle consistency and temporal compositional consistency. We therefore need to explicitly exclude occluded regions in the training loss. Specifically, we compose forward and backward flows, and a pixel is considered valid if it lands close to its original position after forward and then backward mapping. Pixels that violate this consistency are down-weighted through an occlusion mask derived from the residual. Importantly, this procedure evaluates the consistency of coordinate mappings rather than color similarity via reprojection [45]. In practice, we compute this compositional residual directly from the predicted flows, and weight the loss using a robust norm (Eq. (3)), so that unreliable regions are softly suppressed.

Speed. The proposed constraint introduces negligible computational overhead. At 386×496, the interpolation required for flow composition takes approximately 0.00030 seconds, while affine augmentation for both the image and optical flow requires 0.00073 seconds on an NVIDIA RTX Pro 6000 Blackwell GPU, which is negligible in practice. For example, in the setting of Sec. 4.1, when training a RAFT-Large [56] model, the compositional operations and loss evaluation in total account for 0.12% of the total training wall-clock time, indicating that triangular consistency can be incorporated into existing training pipelines with essentially no measurable slowdown.

4 Experiments

To isolate the effect of the proposed compositional constraint, we select two widely used yet relatively minimal optical flow frameworks as baselines. For unsupervised learning, we adopt **ARFlow** [34]. For self-supervised adaptation and supervised learning, we use **RAFT** [56]. These choices allow us to attribute performance changes directly to triangular consistency without auxiliary engineering components. Many subsequent methods, including SMURF [50], largely combine ARFlow-style self-supervised objectives with RAFT-style architectures. Crucially, our method introduces only additional supervision and does not alter the estimator or the original training loss. Thus, it is compatible with existing and future optical flow methods, and improvements observed in our experiments are expected to transfer to subsequent methods.

We test across synthetic and real-world datasets: **FlyingChairs** [11] and **FlyingThings3D** [39] provide large synthetic motion with diverse displacements and occlusions; **MPI-Sintel** [4] offers complex non-rigid motion, and its *Final* setting introduces effects such as motion blur and atmospheric distortions; **KITTI** [14] is a real driving benchmark with ground truth derived from LiDAR/3D reconstruction. We also report zero-shot transfer results on **HD1K** [28] (high-resolution driving sequences) and **Middlebury** [2] (real scenes with small-to-medium motions). We report standard endpoint error (EPE). For KITTI, we report Fl-all, the percentage of optical flow outliers over all ground-truth pixels. A pixel is counted as an outlier when its endpoint error is at least 3 pixels and at least 5% of the ground-truth flow magnitude [41].

4.1 Self-Supervised Adaptation

We first test how triangular consistency *alone* can improve an optical flow model. Our goal is to push the limit of using *only* consistency as supervision, without any auxiliary signals (e.g., photometric reconstruction or smoothness), even if they do not require labels and may result in additional improvement. To this end, we adapt pre-trained optical flow models to new domains without using labeled data in a few-shot setting: a *single* epoch of self-supervised adaptation, after which the adapted model is evaluated on frames that were *not used* during the adaptation stage, resembling a practical test-time adaptation scenario.

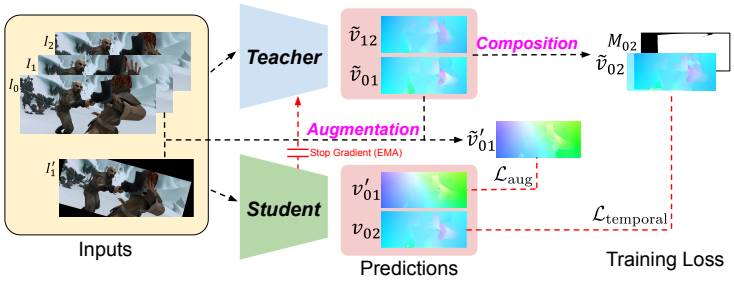

Fig. 2: Self-supervised adaptation with triangular consistency. Given input frames (I_0, I_1, I_2), the teacher network predicts flows $\tilde{v}_{01}$ and $\tilde{v}_{12}$, which are composed to produce the reference flow $\tilde{v}_{02}$. The student network predicts v_{02} and is trained by enforcing consistency with this composed reference. In parallel, a random affine transformation generates an augmented frame $\tilde{I}_1$, inducing the analytically defined flow $\tilde{v}'_{01}$. The student prediction v'_{01} is required to match this reference. The teacher parameters are updated using an exponential moving average (EMA) of the student.

Method. We start from RAFT models pre-trained on FlyingChairs or FlyingThings3D and adapt them to Sintel using neither labels nor photometric or smoothness losses: the only training signal is triangular consistency. Note that we perform adaptation on Sintel's unlabeled test split and evaluate on its labeled training split.[5] This yields a lightweight, test-time-style calibration setting: with batch size 12, one epoch corresponds to 45 iterations and takes 84 seconds on an NVIDIA RTX Pro 6000 Blackwell GPU. Notably, to isolate the effect of the proposed constraint, we freeze normalization statistics during adaptation, so that the model performs no running-stat updates during training (verified by a sanity check). This practice prevents misattributing spurious gains from trivially recalibrating batch statistics to the target data domain.

To stabilize the adaptation process, we adopt a teacher-student self-distillation scheme [16]. The loss is back-propagated only through the student model, while the teacher model is updated using an exponential moving average (EMA) of the student parameters. This scheme is summarized in Fig. 2. The teacher predicts the shorter-range flows $\tilde{v}_{01}$ and $\tilde{v}_{12}$, while the student predicts the longer-range flow v_{02}. We then penalize the discrepancy between the student's direct prediction v_{02} and the composed teacher prediction $\tilde{v}_{02}$. This effectively chains two smaller-displacement, typically more accurate, flows to supervise a larger-displacement flow. In parallel, we enforce augmentation consistency similarly. A random affine transformation is sampled and applied to frame I_1 to produce the augmented frame $\tilde{I}_1$, which also induces flow $\tilde{v}'_{01}$. The loss is computed between the student prediction v'_{01} and the reference flow $\tilde{v}'_{01}$.

[5] Sintel's official evaluation server does not permit repeated evaluation on the test split. We therefore use the available labels from the training split only for evaluation. This experiment is intended as a controlled test of triangular consistency; in practice, the same adaptation can be combined with other forms of self-supervision.

Results. Tab. 1 demonstrates that consistency alone already leads to substantial improvements in optical flow. Even with an extremely small adaptation budget, the pre-trained models consistently improve across Sintel benchmarks by up to 18.1%. These gains are particularly notable given the constrained setting: adaptation is performed using only unlabeled data and relies solely on the proposed consistency objectives, without photometric losses, smoothness losses, or ground-truth labels. Among the proposed components, temporal consistency provides the dominant adaptation signal, improving the error from 2.01 to 1.77 and from 2.54 to 2.18. The augmentation objective further improves these results from 1.77 to 1.73 and from 2.18 to 2.08. In contrast, augmentation-only and cycle-only variants are ineffective, likely because temporal consistency supervises longer-range flow using two shorter, typically more reliable flows, whereas the other terms do not provide a comparable adaptation signal. The fact that such a minimal signal can reliably improve pre-trained models suggests that triangular consistency provides a strong source of supervision for domain adaptation. This is particularly valuable for active learning [71], where high-quality labeled data are scarce and costly to obtain.

Table 1: Self-supervised adaptation by triangular consistency. Under extremely constrained settings: one epoch (45 iterations) using unlabeled data and *only* triangular consistency, accuracy improves significantly.

Pre-training	Method	Clean	Final
FlyingChairs	Pre-trained	2.54	4.67
	+ consistency	**2.08**	**3.95**
	Improvement	18.1%	15.4%
FlyingThings3D	Pre-trained	2.01	3.41
	+ consistency	**1.73**	**3.13**
	Improvement	13.9%	8.2%

4.2 Unsupervised Training

Encouraged by the improvements in the self-supervised adaptation, we now integrate all three forms of triangular consistency into a complete unsupervised optical flow training pipeline.

Method. We build upon ARFlow [34], a widely used unsupervised optical flow framework. ARFlow natively supports loading three consecutive frames during training, allowing triangular consistency to be incorporated without modifying the data loading pipeline. ARFlow primarily relies on photometric reconstruction together with occlusion-aware bidirectional consistency and symmetric augmentation consistency between transformed image pairs as supervision signals. Note that our augmentation consistency differs from ARFlow, since it only augments the target image and is therefore asymmetric. Unlike the self-supervised adaptation experiment, which adapts pre-trained models using unlabeled data, here we train models from scratch using the unlabeled Sintel and KITTI training sets. All training settings and hyperparameters, other than our proposed training loss, follow the publicly available code of ARFlow.

Results. Fig. 3 presents qualitative comparisons on several challenging Sintel sequences. Because triangular consistency constrains the geometry of motion

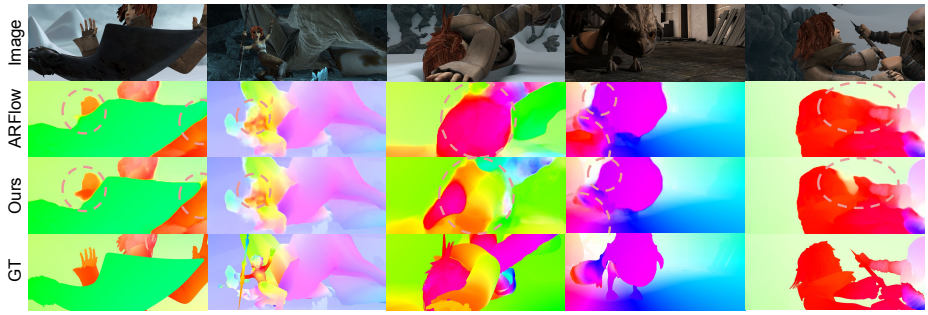

Fig. 3: Qualitative comparison on MPI-Sintel. Regions highlighted by dashed circles illustrate typical improvements introduced by triangular consistency. Our method produces motion fields that better align with the object boundaries and motion continuity of the scene, reflecting the geometric consistency enforced during training.

Table 2: Unsupervised learning with triangular consistency. When training on Sintel, triangular consistency improves both training/test accuracy and cross-dataset performance. When training on KITTI, the training/test error remains nearly unchanged, but zero-shot transfer improves noticeably.

Source	Method	Sintel (train/test) Clean EPE	Sintel (train/test) Final EPE	KITTI (train/test) Fl-all (%)	HD1K EPE	Middlebury EPE
Sintel	ARFlow	2.79 / 4.78	3.73 / 5.89	-	1.40	0.35
	+ ours	**2.58** / **4.48**	**3.49** / **5.86**	-	**1.24**	**0.33**
	Improvement	7.5% / 6.3%	6.4% / 0.5%	-	11.4%	5.7%
KITTI	ARFlow	-	-	**9.87** / 11.80	2.32	0.55
	+ ours	-	-	9.94 / **11.44**	**1.99**	**0.54**
	Improvement	-	-	-0.7% / 3.1%	14.2%	1.8%

correspondences, its effect is visible in the structural coherence of the predicted flow fields. In particular, motion boundaries and large coherent regions exhibit improved alignment with the ground truth. The highlighted regions illustrate typical cases where enforcing consistency across composed flows leads to more stable motion estimates, especially for articulated objects and regions undergoing complex motion. Tab. 2 shows that adding triangular consistency improves unsupervised learning in both *in-domain* accuracy and *cross-dataset* transfer. When training on Sintel, we reduce EPE on both Clean/Final and also improve zero-shot evaluation on HD1K and Middlebury, indicating that compositional and analytic-augmentation constraints provide supervision complementary to photometric reconstruction and bidirectional consistency.

When training on KITTI, metrics on the training set saturate: there is little improvement on KITTI training itself, yet test accuracy still improves by 3.1%. This already suggests that the added constraint primarily improves *generalization*, especially when the training set data distribution is narrow. The benefit becomes more evident when evaluating zero-shot transfer. Models trained with

Table 3: Ablation of triangular consistency losses and loss weights. We report Sintel Clean and Final EPE under unsupervised ARFlow training; lower is better.

Configuration	Clean	Final
Baseline (ARFlow)	2.79	3.73
+ Aug ($\lambda_{\text{aug}} = 0.003$)	2.72	3.70
+ Aug ($\lambda_{\text{aug}} = 0.01$)	2.60	3.56
+ Temp ($\lambda_{\text{temp}} = 0.003$)	2.65	3.68
+ Temp ($\lambda_{\text{temp}} = 0.01$)	2.67	3.69
+ Cyc ($\lambda_{\text{cyc}} = 0.003$)	2.76	3.77
+ Cyc ($\lambda_{\text{cyc}} = 0.005$)	2.69	3.70
+ Cyc ($\lambda_{\text{cyc}} = 0.01$)	2.68	3.69
+ Temp + Cyc ($\lambda_{\text{temp}} = 0.003$, $\lambda_{\text{cyc}} = 0.005$)	2.67	3.68
+ Aug + Temp ($\lambda_{\text{aug}} = 0.01$, $\lambda_{\text{temp}} = 0.003$)	2.63	3.59
+ Aug + Cyc ($\lambda_{\text{aug}} = 0.01$, $\lambda_{\text{cyc}} = 0.005$)	2.61	3.57
+ Aug + Temp + Cyc ($\lambda_{\text{aug}} = 0.02$, $\lambda_{\text{temp}} = 0.003$, $\lambda_{\text{cyc}} = 0.005$)	2.68	3.53
+ Aug + Temp + Cyc ($\lambda_{\text{aug}} = 0.01$, $\lambda_{\text{temp}} = 0.003$, $\lambda_{\text{cyc}} = 0.005$)	**2.58**	**3.49**

triangular consistency achieve substantially better performance on HD1K, reducing EPE by 14.2%. Notably, the same trend is observed in the supervised setting, where triangular consistency introduces motion variations that improve test accuracy and cross-dataset generalization even when the training-set accuracy itself is slightly reduced, reflecting less overfitting.

Ablations. Tab. 3 evaluates the contribution of the three types of triangular consistency. Augmentation consistency gives the largest improvement, reducing the ARFlow baseline from 2.79/3.73 to 2.60/3.56 when $\lambda_{\text{aug}} = 0.01$. Temporal compositional consistency also improves over the baseline, but its gains are smaller and remain similar across the tested weights. Cycle consistency is less stable in isolation: moderate weights improve Clean performance, while Final remains close to the baseline. The combined variants follow the same pattern. Configurations containing augmentation generally outperform temporal and cycle consistency combinations, and the full combination of the three training objectives gives the best overall results, achieving the lowest Clean EPE with $(\lambda_{\text{aug}}, \lambda_{\text{temp}}, \lambda_{\text{cyc}}) = (0.01, 0.003, 0.005)$. This suggests that augmentation consistency is the dominant signal, unlike in self-supervised adaptation where temporal consistency is more effective. A possible reason is that training from scratch and adapting a pre-trained model have different optimization dynamics.

4.3 Supervised Training

Finally, we evaluate triangular consistency under a supervised setting. In this setting, we strictly follow the original RAFT training pipeline [56] and introduce triangular consistency only as a controlled data augmentation mechanism.

Table 4: Triangular consistency as data augmentation for supervised training. We follow the RAFT training pipeline and introduce triangular consistency only through controlled data augmentation. This slightly improves test accuracy and further improves zero-shot transfer across datasets.

Source	Method	Sintel (train/test) Clean EPE	Sintel (train/test) Final EPE	KITTI (train/test) Fl-all (%)	HD1K EPE	Middlebury EPE
Sintel	RAFT	0.76 / 2.08	1.22 / **3.41**	-	0.67	0.29
	+ ours	**0.71** / **2.02**	**1.16** / 3.44	-	**0.66**	**0.27**
	Improvement	6.6% / 2.9%	4.9% / -0.9%	-	1.5%	6.9%
KITTI	RAFT	-	-	**1.53** / 5.27	1.08	0.69
	+ ours	-	-	1.56 / **5.02**	**0.83**	**0.56**
	Improvement	-	-	-2.0% / 4.7%	23.1%	18.8%

Method. We apply random affine transformations to the target image. Using the analytic formulation described in Sec. 3.3, the corresponding pseudo-ground-truth flow is computed accordingly. The model is then trained on the augmented image pair and pseudo-ground-truth. This augmentation effectively preserves exact supervision since the pseudo-ground-truth is computed analytically. Importantly, this is the only modification to the RAFT training pipeline and introduces negligible computational overhead. All other training settings, including model and hyperparameters, remain unchanged.

Results. Tab. 4 shows that triangular consistency consistently improves optical flow under supervised training. We note that Sintel contains large synthetic motions, e.g., action and combat sequences. This induces a bias when transferring to datasets such as Middlebury, which capture motions from natural interactions. By using controlled affine transformations to introduce additional camera-like motion patterns, triangular consistency serves as a regularizer that enables RAFT trained on Sintel to generalize to Middlebury. Hence, despite Sintel having high-fidelity supervision, the proposed triangular consistency still improves cross-dataset generalization by 6.9%. Interestingly, our method does not improve accuracy on the Sintel Final pass, which mainly introduces photometric variations rather than geometric ones. This lack of improvement is consistent with the source of our gains: triangular consistency improves geometric motion generalization, but is not designed to address appearance changes.

The training-set bias is even more pronounced when training on KITTI and testing on HD1K and Middlebury. Since KITTI is an outdoor driving dataset, the motion patterns are largely constrained: forward ego-motion dominates, and viewpoints, motion directions, and displacement magnitudes follow a relatively predictable pattern determined by vehicle dynamics and frame rate. In contrast, Middlebury contains significant motions of people interacting with objects, while the camera motion remains conservative. Training with triangular consistency naturally introduces a much larger set of motion patterns to KITTI. This is reflected in reduced overfitting to the training set and moderate (4.7%) improvement on the KITTI test set, but more substantial improvement when transferring across datasets: 23.1% on HD1K and 18.8% on Middlebury.

This experiment highlights the effectiveness of such a data augmentation mechanism for supervised optical flow. To the best of our knowledge, existing augmentation strategies preserve correspondence by applying identical transformations to both images in a pair [34, 50, 56]. In contrast, we transform only the target image and analytically update the ground-truth. This enables the model to observe a broader set of motion patterns while maintaining exact supervision. Note that, in supervised training, if synthetic data could be generated indefinitely with user-specified motion patterns, such augmentation would be less necessary. In practice, however, synthetic datasets are often released without the simulator, while real-world datasets cannot be resynthesized at scale. In this sense, our method acts as a simulator. The results demonstrate the benefit of such augmentation, particularly under zero-shot transfer when motion statistics are highly constrained in the source dataset.

5 Discussion and Conclusion

5.1 Why was such a simple constraint overlooked?

At first glance, triangular consistency may appear almost self-evident. The compositional nature of displacement fields has been recognized and exploited in tasks involving long-range correspondence [7,30,33]. We were therefore somewhat surprised that such a straightforward geometric relation has not been systematically used for training optical flow. One likely reason is historical: optical flow estimation has primarily been viewed as a pairwise problem [17, 38]. Although multi-frame formulations were explored in classical optimization-based methods, they typically required solving large coupled systems over many frames [13, 19], making them computationally demanding. Modern learning-based optical flow models [11, 54, 56] inherit this pairwise design, and most supervision signals are likewise defined only for pairs of frames.

There are also practical reasons. Existing optical flow training pipelines apply identical geometric transformations to both images to preserve correspondences [34, 50, 56]. In contrast, we augment only the target image and update the ground-truth flow analytically, producing a broader family of motion patterns. Moreover, naively composing flows or applying geometric augmentation often introduces interpolation artifacts or invalid correspondences, particularly near occlusion boundaries and image borders. Our affine formulation alleviates this issue since induced flow can be computed in closed form, even when the transformed image extends beyond the image domain. This avoids interpolation artifacts and allows triangular consistency to be implemented as a stable training signal. Conceptually, this strategy is similar to AugUndo [65], which introduces controlled geometric perturbations to generate additional supervision for depth prediction. As such, we foresee this work extending to other data modalities, e.g., medical images [72, 73], and geometric tasks, e.g., monocular depth estimation [12, 63] and completion [37, 43, 62, 64]. In retrospect, the idea may appear simple, but our experiments show that incorporating this geometric constraint leads to consistent improvements across multiple training regimes.

5.2 Constraints for Optical Flow

Most geometric constraints in optical flow are relatively local. A common example is spatial smoothness, which assumes neighboring pixels move coherently [1, 5, 17, 24, 46]. Some attempts also employ 3D priors as auxiliary supervision signals [9, 25, 44, 76]. Another widely used constraint is forward-backward *symmetry* [21, 34, 49], written as $v_{10} \approx -v_{12}$, which differs from the forward-backward *compositional* consistency we study, written as $w_{10}(w_{01}) \approx \mathrm{Id}$. While effective, it stems from a linear approximation of the trajectory under a temporal-smoothness assumption. Compared to these priors, triangular consistency arises from a more fundamental property: optical flow represents a mapping between coordinate systems, and such mappings compose by construction. Importantly, this relation holds regardless of the specific trajectory taken by the point.

Our compositional constraints do not depend on image appearance and are therefore robust to illumination changes, shadows, reflections, or imaging noise. In this sense, triangular consistency additionally encourages coherent motion estimates across frames, which is evident in Fig. 3. Unlike spatial and temporal smoothness priors, which require users to specify a desired level of smoothness, our proposed constraint does not introduce a smoothness prior. Further, in this work, we focus only on the simplest non-trivial instance of such a compositional structure: a triangle. In principle, these transformations can be chained indefinitely over longer temporal sequences, potentially enabling stronger supervision signals. In future work, this may lead to a curriculum-learning pipeline, where the model is trained with progressively increasing correspondence ranges.

5.3 Limitations

Despite its generality, the effectiveness of triangular consistency depends on the motion statistics of the data. For example, we observe less improvement on the KITTI benchmark, which contains highly restrained motion patterns dominated by forward ego-motion in driving scenes. Consequently, viewpoints, motion directions, and displacement magnitudes follow relatively predictable patterns. In such environments, compositional constraints provide extra supervision, but the benefit is less visible when evaluation follows the original in-domain motion statistics. Importantly, triangular consistency does not hurt performance on KITTI, while models trained on KITTI still demonstrate improvement when tested on HD1K and Middlebury. This observation suggests that geometric consistency constraints are most beneficial when the data contains complex and varied motion that cannot be captured by simple priors.

Another limitation arises from occlusion and multi-layer motion. The compositional relation assumes that a single-layer correspondence exists across frames. In real scenes, occlusions and independently moving objects may violate this assumption. While our implementation mitigates this issue through occlusion masking, incorporating explicit layered motion models or more advanced visibility reasoning may further improve robustness in such scenarios.

5.4 Conclusion

In this work, we revisit a simple yet fundamental property of motion: displacement fields agree under composition. From this intuition, we propose triangular consistency, a minimal compositional constraint that links optical flow estimation across three frames. It depends only on the geometry of motion and therefore remains agnostic to image appearance, network architecture, and supervision type. The resulting method can be integrated into existing optical flow learning pipelines with negligible computational overhead. Despite its simplicity, the proposed mechanism consistently improves optical flow accuracy across unsupervised learning, self-supervised adaptation, and supervised training. In summary, it is fast to compute and functions as a plug-and-play supervision component compatible with existing methods. More broadly, this compositional view is not specific to optical flow: any correspondence problem arising from a temporally discretized continuous process can, in principle, be formulated through consistency under composition.

Acknowledgements

This research was supported by Dong Lao's startup funds at LSU. Computational resources for model training were partially provided by LSU HPC.

We thank Ganesh Sundaramoorthi for bringing up, as early as 2015, the idea of solving long-range correspondence by composition when working on low-latency moving object detection. The composition rule developed at that time later translated into multiple works, which appeared at CVPR 2017, ECCV 2018, ICCV 2021, and CVPR 2024. The idea of controlled geometric augmentation originated from Alex Wong during the development of AugUndo for depth estimation. The initial concept of this work was partially developed at the UCLA Vision Lab during Dong Lao's and Alex Wong's appointments at UCLA.

We would like to give special thanks to the anonymous reviewers for encouraging and appreciating the honest, in-depth discussion of limitations presented in this paper. While demonstrating improvements is important, knowing when and why a method does not improve performance is equally important.

References

1. Bailer, C., Taetz, B., Stricker, D.: Flow fields: Dense correspondence fields for highly accurate large displacement optical flow estimation. In: Proceedings of the IEEE international conference on computer vision. pp. 4015–4023 (2015)
2. Baker, S., Scharstein, D., Lewis, J.P., Roth, S., Black, M.J., Szeliski, R.: A database and evaluation methodology for optical flow. International journal of computer vision **92**(1), 1–31 (2011)
3. Brox, T., Malik, J.: Large displacement optical flow: descriptor matching in variational motion estimation. IEEE transactions on pattern analysis and machine intelligence **33**(3), 500–513 (2010)

4. Butler, D.J., Wulff, J., Stanley, G.B., Black, M.J.: A naturalistic open source movie for optical flow evaluation. In: European conference on computer vision. pp. 611–625. Springer (2012)
5. Chen, Q., Koltun, V.: 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 (2016)
6. Chen, Z., Jin, H., Lin, Z., Cohen, S., Wu, Y.: Large displacement optical flow from nearest neighbor fields. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2443–2450 (2013)
7. Cho, S., Huang, J., Kim, S., Lee, J.Y.: Flowtrack: Revisiting optical flow for long-range dense tracking. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19268–19277 (2024)
8. Doersch, C., Yang, Y., Vecerik, M., Gokay, D., Gupta, A., Aytar, Y., Carreira, J., Zisserman, A.: Tapir: Tracking any point with per-frame initialization and temporal refinement. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10061–10072 (2023)
9. Dong, Q., Cao, C., Fu, Y.: Rethinking optical flow from geometric matching consistent perspective. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition. pp. 1337–1347 (2023)
10. Dong, Q., Fu, Y.: Memflow: Optical flow estimation and prediction with memory. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19068–19078 (2024)
11. Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., Van Der Smagt, P., Cremers, D., Brox, T.: Flownet: Learning optical flow with convolutional networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2758–2766 (2015)
12. Fei, X., Wong, A., Soatto, S.: Geo-supervised visual depth prediction. IEEE Robotics and Automation Letters **4**(2), 1661–1668 (2019)
13. Garg, R., Pizarro, L., Rueckert, D., Agapito, L.: Dense multi-frame optic flow for non-rigid objects using subspace constraints. In: Asian Conference on Computer Vision. pp. 460–473. Springer (2010)
14. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: 2012 IEEE conference on computer vision and pattern recognition. pp. 3354–3361. IEEE (2012)
15. Gibson, J.J.: The perception of the visual world. (1950)
16. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9729–9738 (2020)
17. Horn, B.K., Schunck, B.G.: Determining optical flow. Artificial intelligence **17**(1-3), 185–203 (1981)
18. Huang, Z., Shi, X., Zhang, C., Wang, Q., Cheung, K.C., Qin, H., Dai, J., Li, H.: Flowformer: A transformer architecture for optical flow. In: European conference on computer vision. pp. 668–685. Springer (2022)
19. Irani, M.: Multi-frame optical flow estimation using subspace constraints. In: Proceedings of the Seventh IEEE International Conference on Computer Vision. vol. 1, pp. 626–633. IEEE (1999)
20. Jackson, J.D., Yezzi, A.J., Soatto, S.: Dynamic shape and appearance modeling via moving and deforming layers. International Journal of Computer Vision **79**(1), 71–84 (2008)

21. Janai, J., Guney, F., Ranjan, A., Black, M., Geiger, A.: Unsupervised learning of multi-frame optical flow with occlusions. In: Proceedings of the European conference on computer vision (ECCV). pp. 690–706 (2018)
22. Janai, J., Guney, F., Wulff, J., Black, M.J., Geiger, A.: 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 (2017)
23. Jeong, J., Lin, J.M., Porikli, F., Kwak, N.: Imposing consistency for optical flow estimation. In: Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition. pp. 3181–3191 (2022)
24. Jiang, S., Lu, Y., Li, H., Hartley, R.: Learning optical flow from a few matches. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 16592–16600 (2021)
25. Jiao, Y., Tran, T.D., Shi, G.: Effiscene: Efficient per-pixel rigidity inference for unsupervised joint learning of optical flow, depth, camera pose and motion segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5538–5547 (2021)
26. Jonschkowski, R., Stone, A., Barron, J.T., Gordon, A., Konolige, K., Angelova, A.: What matters in unsupervised optical flow. In: European conference on computer vision. pp. 557–572. Springer (2020)
27. Karaev, N., Rocco, I., Graham, B., Neverova, N., Vedaldi, A., Rupprecht, C.: Cotracker: It is better to track together. In: European conference on computer vision. pp. 18–35. Springer (2024)
28. Kondermann, D., Nair, R., Honauer, K., Krispin, K., Andrulis, J., Brock, A., Gussefeld, B., Rahimimoghaddam, M., Hofmann, S., Brenner, C., et al.: 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 (2016)
29. Lao, D., Sundaramoorthi, G.: Minimum delay moving object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4250–4259 (2017)
30. Lao, D., Sundaramoorthi, G.: Extending layered models to 3d motion. In: Proceedings of the European conference on computer vision (ECCV). pp. 435–451 (2018)
31. Lao, D., Wang, C., Wong, A., Soatto, S.: Diffeomorphic template registration for atmospheric turbulence mitigation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 25107–25116 (2024)
32. Lao, D., Zhu, P., Wonka, P., Sundaramoorthi, G.: Flow-guided video inpainting with scene templates. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 14599–14608 (2021)
33. Le Moing, G., Ponce, J., Schmid, C.: Dense optical tracking: Connecting the dots. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19187–19197 (2024)
34. Liu, L., Zhang, J., He, R., Liu, Y., Wang, Y., Tai, Y., Luo, D., Wang, C., Li, J., Huang, F.: Learning by analogy: Reliable supervision from transformations for unsupervised optical flow estimation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 6489–6498 (2020)
35. Liu, P., King, I., Lyu, M.R., Xu, J.: Ddflow: Learning optical flow with unlabeled data distillation. In: Proceedings of the AAAI conference on artificial intelligence. vol. 33, pp. 8770–8777 (2019)

36. Liu, P., Lyu, M., King, I., Xu, J.: Selflow: Self-supervised learning of optical flow. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)
37. Liu, T.Y., Agrawal, P., Chen, A., Hong, B.W., Wong, A.: Monitored distillation for positive congruent depth completion. In: Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part II. pp. 35–53. Springer (2022)
38. Lucas, B.D., Kanade, T.: 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 (1981)
39. Mayer, N., Ilg, E., Hausser, P., Fischer, P., Cremers, D., Dosovitskiy, A., Brox, T.: 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 (2016)
40. Meister, S., Hur, J., Roth, S.: Unflow: Unsupervised learning of optical flow with a bidirectional census loss. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
41. Menze, M., Geiger, A.: Object scene flow for autonomous vehicles. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3061–3070 (2015)
42. Neoral, M., Šerỳch, J., Matas, J.: Mft: Long-term tracking of every pixel. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 6837–6847 (2024)
43. Park, H., Chen, R., Rim, P., Lao, D., Wong, A.: Orcas: Unsupervised depth completion via occluded region completion as supervision. The Fourteenth International Conference on Learning Representations (2026)
44. Poggi, M., Tosi, F.: Flowseek: optical flow made easier with depth foundation models and motion bases. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5667–5679 (2025)
45. Ren, Z., Yan, J., Ni, B., Liu, B., Yang, X., Zha, H.: Unsupervised deep learning for optical flow estimation. In: Proceedings of the AAAI conference on artificial intelligence. vol. 31 (2017)
46. Revaud, J., Weinzaepfel, P., Harchaoui, Z., Schmid, C.: Epicflow: Edge-preserving interpolation of correspondences for optical flow. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1164–1172 (2015)
47. Ricco, S., Tomasi, C.: Dense lagrangian motion estimation with occlusions. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1800–1807. IEEE (2012)
48. Sand, P., Teller, S.: Particle video: Long-range motion estimation using point trajectories. International journal of computer vision **80**(1), 72–91 (2008)
49. Shi, X., Huang, Z., Bian, W., Li, D., Zhang, M., Cheung, K.C., See, S., Qin, H., Dai, J., Li, H.: Videoflow: Exploiting temporal cues for multi-frame optical flow estimation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12469–12480 (2023)
50. Stone, A., Maurer, D., Ayvaci, A., Angelova, A., Jonschkowski, R.: Smurf: Self-teaching multi-frame unsupervised raft with full-image warping. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3887–3896 (June 2021)
51. Sui, X., Li, S., Geng, X., Wu, Y., Xu, X., Liu, Y., Goh, R., Zhu, H.: Craft: Cross-attentional flow transformer for robust optical flow. In: Proceedings of the

IEEE/CVF conference on Computer Vision and Pattern Recognition. pp. 17602–17611 (2022)
52. Sun, D., Roth, S., Black, M.J.: Secrets of optical flow estimation and their principles. In: 2010 IEEE computer society conference on computer vision and pattern recognition. pp. 2432–2439. IEEE (2010)
53. Sun, D., Sudderth, E.B., Black, M.J.: Layered segmentation and optical flow estimation over time. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1768–1775. IEEE (2012)
54. Sun, D., Yang, X., Liu, M.Y., Kautz, J.: 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 (2018)
55. Sun, S., Chen, Y., Zhu, Y., Guo, G., Li, G.: Skflow: Learning optical flow with super kernels. Advances in Neural Information Processing Systems **35**, 11313–11326 (2022)
56. Teed, Z., Deng, J.: Raft: Recurrent all-pairs field transforms for optical flow. In: European conference on computer vision. pp. 402–419. Springer (2020)
57. Wang, Q., Chang, Y.Y., Cai, R., Li, Z., Hariharan, B., Holynski, A., Snavely, N.: Tracking everything everywhere all at once. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 19795–19806 (2023)
58. Wang, X., Jabri, A., Efros, A.A.: Learning correspondence from the cycle-consistency of time. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2566–2576 (2019)
59. Wang, Y., Yang, Y., Yang, Z., Zhao, L., Wang, P., Xu, W.: Occlusion aware unsupervised learning of optical flow. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4884–4893 (2018)
60. Wang, Y., Lipson, L., Deng, J.: Sea-raft: Simple, efficient, accurate raft for optical flow. In: European Conference on Computer Vision. pp. 36–54. Springer (2024)
61. Wen, H., Liang, E., Deng, J.: Layeredflow: A real-world benchmark for non-lambertian multi-layer optical flow. In: European Conference on Computer Vision. pp. 477–495. Springer (2024)
62. Wong, A., Fei, X., Tsuei, S., Soatto, S.: Unsupervised depth completion from visual inertial odometry. IEEE Robotics and Automation Letters **5**(2), 1899–1906 (2020)
63. Wong, A., Soatto, S.: Bilateral cyclic constraint and adaptive regularization for unsupervised monocular depth prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5644–5653 (2019)
64. Wong, A., Soatto, S.: Unsupervised depth completion with calibrated backprojection layers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12747–12756 (2021)
65. Wu, Y., Liu, T.Y., Park, H., Soatto, S., Lao, D., Wong, A.: Augundo: Scaling up augmentations for monocular depth completion and estimation. In: European Conference on Computer Vision. pp. 274–293. Springer (2024)
66. Xu, L., Jia, J., Matsushita, Y.: Motion detail preserving optical flow estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence **34**(9), 1744–1757 (2011)
67. Yang, Y., Lu, Z., Sundaramoorthi, G.: Coarse-to-fine region selection and matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5051–5059 (2015)
68. Yang, Y., Soatto, S.: S2f: Slow-to-fast interpolator flow. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2087–2096 (2017)

69. Yang, Y., Soatto, S.: Conditional prior networks for optical flow. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 271–287 (2018)
70. Yu, J.J., Harley, A.W., Derpanis, K.G.: Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness. In: European conference on computer vision. pp. 3–10. Springer (2016)
71. Yuan, S., Sun, X., Kim, H., Yu, S., Tomasi, C.: Optical flow training under limited label budget via active learning. ArXiv **abs/2203.05053** (2022), `https://api.semanticscholar.org/CorpusID:247362834`
72. Zhang, X., Pak, D.H., Ahn, S.S., Li, X., You, C., Staib, L.H., Sinusas, A.J., Wong, A., Duncan, J.S.: Heteroscedastic uncertainty estimation framework for unsupervised registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 651–661. Springer (2024)
73. Zhang, X., Stendahl, J.C., Staib, L.H., Sinusas, A.J., Wong, A., Duncan, J.S.: Adaptive correspondence scoring for unsupervised medical image registration. In: European Conference on Computer Vision. pp. 76–92. Springer (2024)
74. Zhao, S., Zhao, L., Zhang, Z., Zhou, E., Metaxas, D.: Global matching with overlapping attention for optical flow estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17592–17601 (2022)
75. Zhou, T., Jae Lee, Y., Yu, S.X., Efros, A.A.: Flowweb: Joint image set alignment by weaving consistent, pixel-wise correspondences. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1191–1200 (2015)
76. Zou, Y., Luo, Z., Huang, J.B.: Df-net: Unsupervised joint learning of depth and flow using cross-task consistency. In: Proceedings of the European conference on computer vision (ECCV). pp. 36–53 (2018)