U^{2}Flow: Uncertainty-Aware Unsupervised Optical Flow Estimation

Xunpei Sun[1], Wenwei Lin[1], Yi Chang[2], Gang Chen[1*]
[1]Sun Yat-sen University, Guangzhou, China
[2]Huazhong University of Science and Technology, Wuhan, China
sunxp7@mail2.sysu.edu.cn, linww28@mail.sysu.edu.cn, yichang@hust.edu.cn,
cheng83@mail.sysu.edu.cn

Abstract

Unsupervised optical flow methods typically lack reliable uncertainty estimation, limiting their robustness and interpretability. We propose U^{2}Flow, the first recurrent unsupervised framework that jointly estimates optical flow and per-pixel uncertainty. The core innovation is a decoupled learning strategy that derives uncertainty supervision from augmentation consistency via a Laplace-based maximum likelihood objective, enabling stable training without ground truth. The predicted uncertainty is further integrated into the network to guide adaptive flow refinement and dynamically modulate the regional smoothness loss. Furthermore, we introduce an uncertainty-guided bidirectional flow fusion mechanism that enhances robustness in challenging regions. Extensive experiments on KITTI and Sintel demonstrate that U^{2}Flow achieves state-of-the-art performance among unsupervised methods while producing highly reliable uncertainty maps, validating the effectiveness of our joint estimation paradigm. The code is available at https://github.com/sunzunyi/U2FLOW.

1. Introduction

Optical flow estimation [6, 42, 43] is a fundamental vision task with broad applications [17, 22, 36, 58]. Recently, deep recurrent models based on all-pairs correlation, such as RAFT [5, 11, 28, 38, 39, 45, 47], have achieved state-of-the-art performance in fully supervised settings.

However, obtaining large-scale, pixel-accurate ground-truth optical flow is costly and often impractical [4, 10, 44, 53], motivating research on unsupervised and self-supervised optical flow estimation [24–26]. Nevertheless, due to the absence of reliable supervision, self-supervised models often produce inaccurate estimates, especially when facing intrinsic challenges such as occlusions [29, 32, 50], textureless regions [15, 31], and large motion displacements [46]. These estimation errors can be detrimental to downstream tasks. In point tracking [21] and multi-view reconstruction [34], even small local optical flow errors can accumulate and lead to failure. Moreover, erroneous flow estimates often cause artifacts in depth recovery [27, 59] or motion-based segmentation [56]. Consequently, it is not enough for a model to predict what the motion is; it must also quantify how confident it is in that prediction [51].

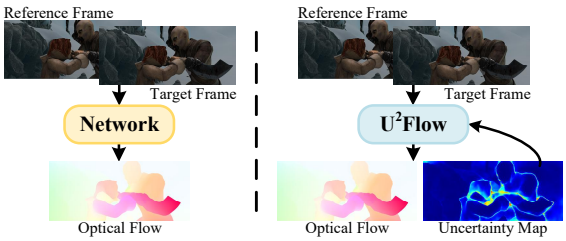


Figure 1. Comparison between previous optical flow estimation methods and our approach. (Left) Previous methods estimate only optical flow. (Right) Our proposed **U^{2}Flow** framework jointly estimates optical flow and its uncertainty, and further leverages the predicted uncertainty to refine the flow estimation.

Despite its importance, uncertainty estimation in the self-supervised setting remains largely underexplored, primarily due to two core challenges: 1) The Absence of Direct Supervision: Unlike supervised methods [12, 13, 16, 37], which can be trained with ground-truth variance or likelihood information, self-supervised models lack any explicit "correct answer" for uncertainty. A fundamental difficulty lies in teaching a model to assess its own reliability without access to ground truth. 2) The Effective Integration of Uncertainty: Even if uncertainty can be estimated, it is unclear how to leverage it effectively during training to improve flow accuracy, rather than treating it as a mere byproduct.

To address these challenges, we present **U^{2}Flow**, the first recurrent framework for the self-supervised joint estimation of dense optical flow and its per-pixel uncertainty, as illustrated in Fig. 1. First, to overcome the lack of direct uncertainty supervision, we derive a supervisory signal from the model's own predictive inconsistencies under data augmentation. When the model yields inconsistent predictions

*Corresponding author.

under diverse spatial and appearance perturbations, it exposes its own regions of low confidence, thereby providing a powerful self-supervisory signal for uncertainty learning. By enforcing consistency between flows predicted under various perturbations, we derive a Laplace-based maximum likelihood objective [8, 12] that learns uncertainty distributions, decoupled from the main flow loss.

Second, to effectively leverage the predicted uncertainty during training, we design an uncertainty-aware recurrent architecture. The predicted uncertainty is fed back into the recurrent update block to guide adaptive refinement. Concurrently, uncertainty is used to modulate the self-supervision objectives, enabling the model to intelligently down-weight unreliable signals.

Experiments show that U^{2}Flow achieves state-of-the-art performance among unsupervised methods on the KITTI and Sintel benchmarks. Crucially, it also produces highly reliable uncertainty maps, demonstrating the efficacy of our joint estimation framework.

Our main contributions are summarized as follows:

- To the best of our knowledge, we propose the first recurrent unsupervised framework for jointly estimating optical flow and uncertainty, which integrates an uncertainty prediction head and an uncertainty-aware refinement mechanism within the recurrent update block.
- We devise a decoupled uncertainty learning strategy based on augmentation consistency for stable estimation. Furthermore, we design an uncertainty-guided regional smoothness mechanism that leverages confidence to improve flow coherence.
- We propose an uncertainty-guided bidirectional flow fusion mechanism that utilizes uncertainty from both forward and backward flows to correct unreliable regions, outperforming traditional occlusion-based strategies.

2. Related Work

Unsupervised Optical Flow: Unsupervised optical flow estimation typically relies on photometric consistency losses combined with spatial smoothness regularization [15, 29, 32, 35]. However, the photometric signal often becomes unreliable in the presence of occlusions, motion blur, illumination changes, or textureless regions [25, 26, 31, 46, 50]. Early works [32, 50] explicitly handled occlusions by deriving binary masks to remove geometrically inconsistent pixels from the photometric loss.

Subsequent approaches, such as ARFlow [24] and SelFlow [26], improved robustness through knowledge distillation and extensive data augmentation [25, 54, 55]. More recent efforts, including SemARFlow [54] and UnSAMFlow [55], incorporated semantic cues to enhance motion boundary preservation. Other methods [14, 41, 46] leverage multi-frame training to provide richer temporal context and complementary motion evidence.

However, most models focus solely on point estimation and ignore prediction uncertainty, limiting their ability to differentiate between ambiguous and reliable regions.

Uncertainty Estimation: Quantifying model confidence is crucial for building robust computer vision systems [1, 23, 49, 57]. Early optical flow approaches typically treated uncertainty as a post-processing step [2, 19, 20, 30], estimating confidence heuristically from image gradients [2] or local flow energy [3] rather than integrating it into the learning process. Subsequent works introduced probabilistic formulations for joint estimation of optical flow and uncertainty [7, 8, 12, 51, 52]. ProbFlow [51] uses variational inference within a probabilistic framework to jointly estimate flow and uncertainty. PDC-Net+ [48] jointly learns dense correspondences and their associated uncertainties under supervised synthetic data, while ProbDiffFlow [60] outputs multiple flow hypotheses instead of a single result. Abdein et al. [1] map the flow smoothness error to a probability distribution to model uncertainty.

Crucially, most existing optical flow uncertainty methods depend on full supervision [12, 13, 16, 37], tightly coupling uncertainty learning with flow regression. This dependency makes them incompatible with the self-supervised paradigm where no such ground truth is available. Moreover, uncertainty is rarely leveraged to improve the flow estimation process itself during inference.

3. Method

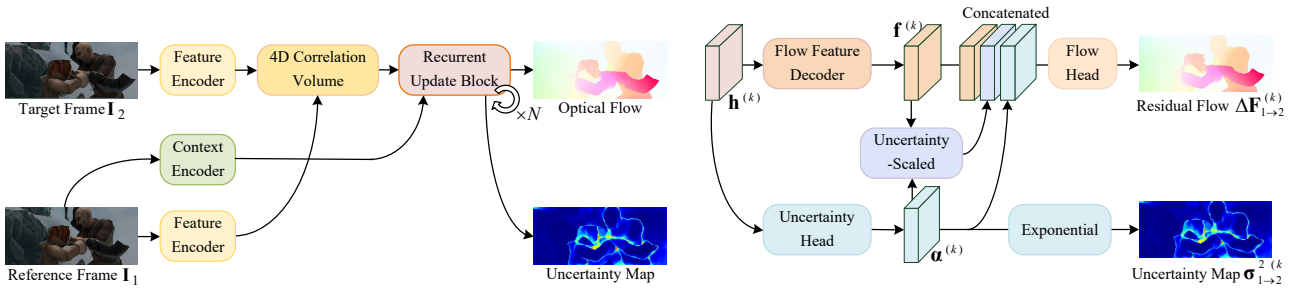
Given two consecutive RGB frames $\mathbf{I}_1, \mathbf{I}_2 \in \mathbb{R}^{H\times W\times 3}$, our goal is to estimate the dense optical flow field $\mathbf{F}_{1\to 2} \in \mathbb{R}^{H\times W\times 2}$ along with its corresponding uncertainty map $\boldsymbol{\sigma}^2_{1\to 2} \in \mathbb{R}^{H\times W}$.

3.1. Network Overview

U^{2}Flow inherits the core design of RAFT [47], as illustrated in Fig. 2a. Given two consecutive frames $\mathbf{I}_1, \mathbf{I}_2$, we first extract deep feature representations $\mathbf{f}_1, \mathbf{f}_2 \in \mathbb{R}^{H'\times W'\times D}$ and construct a 4D correlation volume $\mathbf{C} \in \mathbb{R}^{H'\times W'\times H'\times W'}$, which encodes pairwise similarities between all feature locations. This correlation volume is iteratively queried by a recurrent update block that refines the optical flow estimate $\mathbf{F}^{(k)}_{1\to 2}$ through multiple iterations $k = 1, 2, \ldots, K$.

Following SMURF [41], we replace all batch normalization layers with instance normalization to enhance training stability and convergence under small-batch, unsupervised settings. Furthermore, U^{2}Flow introduces an additional uncertainty head to estimate the flow uncertainty $\boldsymbol{\sigma}^{2\,(k)}_{1\to 2}$ at each iteration. This naturally raises an important question: can the estimated uncertainty be utilized within the network itself to further refine the optical flow?

To address this, we reformulate the original flow head into an uncertainty-aware refinement module that leverages



(a) Overall network structure

(b) Flow refinement module and uncertainty estimation head

Figure 2. Overview of **U^{2}Flow** architecture. (a) The overall recurrent structure follows RAFT [47]. (b) The uncertainty-aware refinement module and the uncertainty estimation head respectively predict optical flow and per-pixel uncertainty in the recurrent update block.

the predicted uncertainty to guide the iterative flow refinement process (see Sec. 3.2). Meanwhile, during training, the refined flow supervises the uncertainty branch, progressively enhancing its estimation accuracy (see Sec. 3.3). In this way, U^{2}Flow jointly learns optical flow and uncertainty within a unified optimization framework.

3.2. Uncertainty Estimation and Flow Refinement

As depicted in Fig. 2b, after obtaining the hidden representation from the RAFT recurrent unit, U^{2}Flow first applies an uncertainty estimation head and then feeds its output into our refinement module. Given the hidden feature representation $\mathbf{h}^{(k)}$ at iteration k, we first decode an intermediate flow feature and estimate its corresponding uncertainty as

$$\mathbf{f}^{(k)} = \mathcal{C}_{\text{flow}}(\mathbf{h}^{(k)}), \qquad \boldsymbol{\alpha}^{(k)} = \mathcal{C}_{\text{unc}}(\mathbf{h}^{(k)}), \tag{1}$$

where $\mathcal{C}_{\text{flow}}(\cdot)$ and $\mathcal{C}_{\text{unc}}(\cdot)$ denote convolutional layers for flow feature extraction and uncertainty estimation, respectively. To ensure strictly positive uncertainty values, the uncertainty head predicts the logarithm of the flow variance:

$$\boldsymbol{\alpha}^{(k)} = \log\left(\boldsymbol{\sigma}^{2\,(k)}_{1\to2}\right), \tag{2}$$

which improves numerical stability during training. During inference, the actual flow uncertainty (i.e., the variance) is recovered as $\boldsymbol{\sigma}^{2\,(k)}_{1\to2} = \exp(\boldsymbol{\alpha}^{(k)})$.

To modulate the influence of uncertain regions, an uncertainty weight map is obtained via a sigmoid transformation:

$$\mathbf{s}^{(k)} = \phi(-\boldsymbol{\alpha}^{(k)}), \tag{3}$$

which acts as a reliability indicator for each flow vector. The uncertainty-scaled flow feature is defined as

$$\tilde{\mathbf{f}}^{(k)} = \mathbf{f}^{(k)} \odot \mathbf{s}^{(k)*}, \tag{4}$$

where $\odot$ denotes element-wise multiplication, and $(\cdot)^*$ represents a stop-gradient operation that prevents backpropagation through the uncertainty branch.

Finally, the refined optical flow residual is obtained by fusing the original feature, the scaled feature, and the uncertainty map:

$$\Delta\mathbf{F}^{(k)}_{1\to2} = \mathcal{C}'_{\text{flow}}\left(\text{concat}\left(\mathbf{f}^{(k)}, \tilde{\mathbf{f}}^{(k)}, \boldsymbol{\alpha}^{(k)*}\right)\right), \tag{5}$$

where $\mathcal{C}'_{\text{flow}}(\cdot)$ denotes the convolutional head that outputs the flow residual $\Delta\mathbf{F}^{(k)}_{1\to2} \in \mathbb{R}^{H'\times W'\times 2}$.

This uncertainty-aware refinement mechanism enables the model to dynamically use predicted uncertainty to modulate flow features, effectively suppressing the influence of unreliable regions during refinement.

3.3. Uncertainty-Aware Unsupervised Loss

Unlike standard unsupervised optical flow methods, our loss explicitly integrates predicted uncertainty to modulate the training signal, enabling joint flow and uncertainty estimation within a purely self-supervised framework.

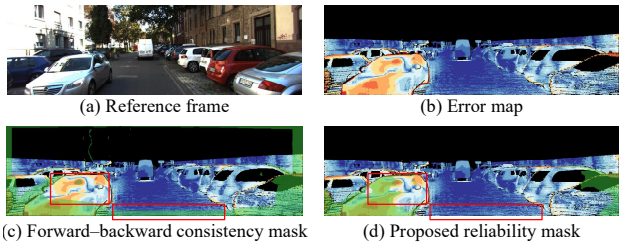
Photometric Loss: During each refinement iteration k, the input frames are warped by $\mathbf{F}^{(k)}_{1\to2}$ and $\mathbf{F}^{(k)}_{2\to1}$ to synthesize a new view of the source frame. Similar to ARFlow [24], the photometric loss $\ell^{(k)}_{\text{ph}}$ between each original image and its warped counterpart is computed as a weighted combination of three terms: the pixel-wise ℓ_1 distance, SSIM, and the census loss [32]. The occlusion mask $\mathbf{O}^{(k)}_{i\to j}$ is computed using a forward–backward consistency check [32] to exclude regions without valid correspondences:

$$\mathbf{O}^{(k)}_{i\to j}(p) = \mathbb{1}\left(\left\|\mathbf{F}^{(k)}_{i\to j}(p) + \mathbf{F}^{(k)}_{j\to i}(p + \mathbf{F}^{(k)}_{i\to j}(p))\right\|_2 > \delta\right), \tag{6}$$

where δ represents an adaptive threshold, $\mathbb{1}(\cdot)$ is the indicator function, $(i,j) \in \{(1,2),(2,1)\}$ refers to the flow direction, and p indicates the pixel coordinates.

Smoothness Loss: To encourage locally coherent flow while preserving motion boundaries, we apply an edge-aware smoothness regularization to the predicted flow $\mathbf{F}^{(k)}_{i\to j}$ at each iteration k, denoted as $\ell^{(k)}_{\text{sm}}$. Building on the work of UnSAMFlow [55], we also incorporate a regional smoothness constraint based on homography estimation. For each object region, a homography is estimated via RANSAC from reliable correspondences derived from the current flow prediction. The resulting homographies are then used to generate regionally refined flow fields.

A key distinction of our approach lies in how reliable correspondences are identified. While the original formulation employs occlusion masks to exclude unreliable pixels,



(a) Reference frame (b) Error map

(c) Forward–backward consistency mask (d) Proposed reliability mask

Figure 3. Comparison of different masks for indicating flow errors. The masks are visualized as translucent green overlays on the optical flow error maps, where correct estimations are shown in blue, incorrect ones in red, and black pixels denote regions without ground truth. As shown in (c), some high-error regions remain unmasked, while certain low-error regions are incorrectly masked. In contrast, our method (d) accurately identifies high-error regions, providing more reliable cues for subsequent homography smoothness (Sec. 3.3), flow fusion (Sec. 3.4), and other downstream tasks.

we introduce a more principled uncertainty-based reliability mask. This mechanism leverages the model's predicted uncertainty, excluding pixels with uncertainty above a threshold τ_{hg} (see Fig. 3). It ensures that only high-confidence regions contribute to homography estimation and optimization. Consequently, the smoothness regularization becomes adaptive, directly guided by the model's learned confidence. It is worth noting that due to the stringent planarity assumption of homography, this uncertainty-enhanced component is applied exclusively to the KITTI [9, 33] dataset, which contains predominantly planar rigid motions. The effectiveness of this uncertainty-guided strategy is further validated in our ablation studies (Sec. 4.5).

The homography smoothness loss is defined as the ℓ_1 distance between the predicted flow $\mathbf{F}_{i\to j}$ and the homography-refined flow $\mathbf{F}^{\text{H}}_{i\to j}$:

$$\ell_{\text{hg}} = \frac{1}{H' \times W'} \sum\nolimits_{\boldsymbol{p}} \left\| \mathbf{F}_{i\to j}(\boldsymbol{p}) - \mathbf{F}^{\text{H}}_{i\to j}(\boldsymbol{p}) \right\|_1 . \tag{7}$$

Augmentation and Uncertainty Supervision: In the absence of ground-truth, we generate a supervisory signal for the uncertainty head through the principle of augmentation consistency. The process begins by computing an initial flow estimate, $\mathbf{F}_{1\to 2}$, for an image pair $(\mathbf{I}_1, \mathbf{I}_2)$ in an initial forward pass. We then apply a set of strong appearance and spatial augmentations to both the images and the flow field, producing an augmented pair $(\hat{\mathbf{I}}_1, \hat{\mathbf{I}}_2)$ and a transformed pseudo-ground-truth flow $\hat{\mathbf{F}}_{1\to 2}$. The network then re-estimates the flow on the augmented pair, yielding a new prediction $\hat{\mathbf{F}}'_{1\to 2}$. The ℓ_1 distance between these flows, $\hat{D}^{(k)}(\boldsymbol{p}) = \|\hat{\mathbf{F}}_{1\to 2}(\boldsymbol{p}) - \hat{\mathbf{F}}'^{(k)}_{1\to 2}(\boldsymbol{p})\|_1$, serves as the self-supervised target, as it captures the model's predictive inconsistency under perturbation.

This augmentation consistency introduces diverse appearance perturbations (e.g., color jitter, contrast adjustment, Gaussian noise, random erasing) and spatial transformations (e.g., translation, random rotation, rescaling), thereby exposing the network to a wide range of uncertainty conditions. To capture this uncertainty, we adopt a Maximum Likelihood Estimation (MLE) formulation. For each augmentation iteration k, the objective is to minimize the negative log-likelihood (NLL) of observing the target flow $\hat{\mathbf{F}}_{1\to 2}$ given the predicted distribution:

$$\mathcal{L}^{(k)}_{\text{unc}} = -\mathbb{E}_{\boldsymbol{p}} \left[\log p\Big(\hat{\mathbf{F}}_{1\to 2}(\boldsymbol{p}) \mid \hat{\mathbf{F}}'^{(k)}_{1\to 2}(\boldsymbol{p}), \boldsymbol{\sigma}^{2(k)}(\boldsymbol{p}) \Big) \right] . \tag{8}$$

Following [12], we assume a Laplace likelihood, which aligns naturally with the ℓ_1-based residual $\hat{D}^{(k)}$. Under the independence assumption across flow dimensions, this leads to the numerically stable MLE objective:

$$\begin{cases} \tilde{\ell}^{(k)}_{\text{unc}}(\boldsymbol{p}) = \sqrt{2}\, \exp\!\big(-\tfrac{1}{2}\boldsymbol{\alpha}^{(k)}(\boldsymbol{p})\big)\, \hat{D}^{(k)}(\boldsymbol{p}) + \tfrac{1}{2}\boldsymbol{\alpha}^{(k)}(\boldsymbol{p}), \\ \ell^{(k)}_{\text{unc}} = \dfrac{\sum_{\boldsymbol{p}} (1 - \hat{O}_{1\to 2}(\boldsymbol{p}))\, \tilde{\ell}^{(k)}_{\text{unc}}(\boldsymbol{p})}{\sum_{\boldsymbol{p}} (1 - \hat{O}_{1\to 2}(\boldsymbol{p}))} . \end{cases} \tag{9}$$

Here, $\boldsymbol{\alpha}^{(k)} = \log \boldsymbol{\sigma}^{2(k)}$ denotes the predicted log-variance, and $\hat{O}_{1\to 2}$ is the transformed occlusion map [24]. Unlike prior works [12, 13, 16, 37] that tightly couple flow regression and uncertainty estimation via a single MLE objective, our framework adopts a decoupled design, separating the flow loss from the uncertainty loss. Consequently, we detach $\hat{D}^{(k)}$ from the MLE objective to prevent gradient leakage into the main flow estimation branch, which improves the robustness and stability of self-supervised learning (see Sec. 4.5 for ablations).

To ensure the flow estimation branch continues to benefit from augmentation regularization, we retain the standard augmentation loss $\ell^{(k)}_{\text{ar}}$, which operates on the same residual $\hat{D}^{(k)}$ with gradient flow enabled, and further incorporate semantic augmentations [55]. The augmentation regularization loss is defined as the ℓ_1 distance between the transformed and predicted flows:

$$\ell^{(k)}_{\text{ar}} = \frac{\sum_{\boldsymbol{p}} (1 - \hat{O}_{1\to 2}(\boldsymbol{p}))\, \hat{D}^{(k)}(\boldsymbol{p})}{\sum_{\boldsymbol{p}} (1 - \hat{O}_{1\to 2}(\boldsymbol{p}))} . \tag{10}$$

A similar formulation applies for semantic augmentations, denoted as $\ell^{(k)}_{\text{sem}}$ [55].

Final Loss: The overall training objective integrates all loss components as follows:

$$\begin{aligned} \ell_{\text{Total}} = \lambda_{\text{hg}}\, \ell_{\text{hg}} + \sum_{k=1}^{K} \zeta^{K-k} \Big(& \ell^{(k)}_{\text{ph}} + \lambda_{\text{sm}}\, \ell^{(k)}_{\text{sm}} + \lambda_{\text{ar}}\, \ell^{(k)}_{\text{ar}} \\ & + \lambda_{\text{sem}}\, \ell^{(k)}_{\text{sem}} + \lambda_{\text{unc}}\, \ell^{(k)}_{\text{unc}} \Big), \end{aligned} \tag{11}$$

where ζ is an exponential decay factor that assigns smaller weights to earlier iterations. The homography smoothness term ℓ_{hg} is applied only to the final iteration output.

	Method	Sintel Clean		Sintel Final		KITTI 2012				KITTI 2015		
		train	test	train	test	train	test			train	test	
		EPE	EPE	EPE	EPE	EPE	Fl-all	Fl-noc	EPE	EPE	Fl-all	Fl-noc
Supervised	PWC-Net+ [43]	(1.71)	3.45	(2.34)	4.60	(0.99)	6.72	3.36	1.4	(1.47)	7.72	4.91
	RAFT [47]	(0.77)	1.61	(1.27)	2.86	–	–	–	–	(0.63)	5.10	3.07
	FlowFormer [11]	(0.48)	1.16	(0.74)	2.09	–	–	–	–	(0.53)	4.68	2.69
	VideoFlow [38]^{MF}	(0.46)	**0.99**	(0.66)	**1.62**	–	–	–	–	(0.56)	**3.65**	–
	FlowDiffuser [28]	–	1.02	–	2.03	–	–	–	–	–	4.17	2.82
Unsupervised	UnFlow-CSS [32]	–	9.38	(7.91)	10.22	3.29	–	–	–	8.10	23.27	–
	SelFlow [26]^{MF}	(2.88)	6.56	(3.87)	6.57	1.69	7.68	4.31	2.2	4.84	14.19	9.65
	UFlow [15]	(2.50)	5.21	(3.39)	6.50	1.68	7.91	4.26	1.9	2.71	11.13	8.41
	ARFlow [24]	(2.79)	4.78	(3.73)	5.89	1.44	–	–	1.8	2.85	11.80	–
	UPFlow [29]	(2.33)	4.68	(2.67)	5.32	1.27	–	–	1.4	2.45	9.38	–
	SMURF [41]^{MF}	(1.71)	3.15	(2.58)	4.18	–	**6.19**	**3.13**	1.4	2.00	6.83	5.26
	SemARFlow [54]†	–	–	–	–	1.28	7.35	3.90	1.5	2.18	8.38	5.43
	UnSAMFlow [55]†	(2.21)	3.93	(3.07)	5.20	1.26	7.05	3.79	1.4	2.01	7.83	5.67
	M2Flow [46]^{MF}	(2.01)	3.38	(3.12)	5.01	**1.09**	6.24	3.95	**1.2**	1.95	7.37	5.73
	U^{2}Flow (Ours)	(1.42)	2.83	(2.32)	4.16	1.19	6.37	3.48	1.4	1.83	6.13	4.56
	U^{2}Flow (Ours +FF)^{MF}	(**1.36**)	**2.83**	(**2.29**)	**4.10**	1.12	6.26	3.47	1.3	**1.74**	**6.00**	**4.52**

Table 1. Quantitative results on Sintel and KITTI online benchmarks. Metrics evaluated at "all" (all pixels), "noc" (non-occlusions). *MF* denotes methods trained using multi-frame data. † denotes models with semantic inputs. "+FF" denotes our bidirectional flow fusion module. Missing entries (–) denote unreported results. Parentheses indicate that training and testing are conducted on the same dataset.

3.4. Uncertainty-Guided Bidirectional Flow Fusion

We leverage the predicted uncertainty to guide a bidirectional flow fusion process, enhancing the reliability of the final estimate. Inspired by SMURF [41], we employ a lightweight convolutional network to learn a mapping from the backward flow $\mathbf{F}_{t\to t-1}$ to the forward flow $\mathbf{F}_{t\to t+1}$. However, our approach fundamentally departs from prior work in its supervisory signal. Instead of relying on heuristic occlusion masks, we train the fusion network exclusively in high-confidence regions identified by our learned uncertainty. Specifically, a region is deemed high-confidence only if both its forward and backward uncertainty estimates are below a given threshold θ:

$$\mathbf{M}_{\text{f}} = \mathbb{1}(\boldsymbol{\sigma}^2_{t\to t+1} < \theta), \quad \mathbf{M}_{\text{b}} = \mathbb{1}(\boldsymbol{\sigma}^2_{t\to t-1} < \theta), \quad (12)$$

where $\mathbf{M}_{\text{f}}$ and $\mathbf{M}_{\text{b}}$ are the resulting reliability masks. During inference, this trained network corrects the initial forward flow. The final fused flow $\mathbf{F}^{\text{fused}}_{t\to t+1}$ is computed by adaptively fusing the original estimate with the prediction $\bar{\mathbf{F}}_{t\to t+1}$, which is derived from $\mathbf{F}_{t\to t-1}$ via the lightweight convolutional network:

$$\begin{cases} \mathbf{F}^{\text{fused}}_{t\to t+1} = \mathbf{F}_{t\to t+1} \odot (1 - \mathbf{M}_{\text{fused}}) + \bar{\mathbf{F}}_{t\to t+1} \odot \mathbf{M}_{\text{fused}}, \\ \mathbf{M}_{\text{fused}} = (1 - \mathbf{M}_{\text{f}}) \odot \mathbf{M}_{\text{b}}. \end{cases} \quad (13)$$

The fusion mask $\mathbf{M}_{\text{fused}}$ activates this correction precisely where the forward flow is uncertain ($\mathbf{M}_{\text{f}} = 0$) but the backward flow is confident ($\mathbf{M}_{\text{b}} = 1$).

This uncertainty-based fusion strategy fundamentally differentiates our approach from methods like SMURF, which primarily focus on correcting flow within occluded regions (often identified by forward-backward consistency checks [32]). However, true occlusion does not always equate to poor flow estimation, and conversely, non-occluded regions can still yield highly unreliable flow predictions, as illustrated in Fig. 3. By leveraging the reliability masks $\mathbf{M}_{\text{f}}$ and $\mathbf{M}_{\text{b}}$, our method's correction mechanism extends beyond strict occlusion handling. Furthermore, this fusion operates as a lightweight refinement on U^{2}Flow outputs, achieving multi-frame benefits without the extensive retraining on large-scale datasets required by methods like SMURF [41]. (See fusion ablation results in Sec. 4.5.)

4. Experiments

4.1. Implementation Details

Dataset: To ensure a fair comparison, we evaluate our method on the KITTI [9, 33] and Sintel [4] datasets, following the training data schedule of prior works [26, 46, 55].
Training: Our model is trained using Adam [18] ($\beta_1 = 0.9$, $\beta_2 = 0.999$) with a batch size of 4. The training first runs for 100k iterations on raw data with a fixed learning rate of 2×10^{-4}, followed by fine-tuning on the original dataset using the OneCycleLR scheduler [40] with a maximum learning rate of 2.5×10^{-4} (100k iterations for KITTI and 50k for Sintel). Augmentation regular-

Method	Sintel		KITTI	
	AUSE ↓	CC ↑	AUSE ↓	CC ↑
FB Check [32]	0.19	0.57	0.21	0.53
PDC-Net+ [48]	0.18	0.45	0.16	0.50
Smoothness [1]	0.26	0.43	0.23	0.45
U^{2}Flow (Ours)	**0.11**	**0.66**	**0.12**	**0.64**

Table 2. Comparison of uncertainty estimation performance on the Sintel (final, clean) and KITTI (2012, 2015) training sets.

ization (appearance and spatial transformations) is introduced after 50k iterations, while the edge-aware smoothness term ℓ_{sm} is deactivated. The homography smoothness loss ℓ_{hg} and semantic augmentation loss ℓ_{sem} are activated after 50% of the fine-tuning iterations. The hyperparameters are set to $K = 12$, $[\lambda_{\text{hg}}, \lambda_{\text{sm}}, \lambda_{\text{ar}}, \lambda_{\text{sem}}, \lambda_{\text{unc}}, \tau_{\text{hg}}] = [0.1, 55, 0.02, 0.05, 0.005, 2]$, and $\theta = 45$ for Sintel and 35 for KITTI.

For data augmentation, we follow ARFlow [24], applying appearance transformations (brightness, contrast, saturation, hue, Gaussian blur, *etc.*), as well as random flipping and swapping. All input images are resized to 256×832 for KITTI and 448×1024 for Sintel.

4.2. Benchmark Testing

We evaluate our method using standard optical flow metrics, including the average endpoint error (EPE) and the percentage of erroneous pixels (Fl). Comparisons are conducted against both supervised and unsupervised approaches on the KITTI and Sintel benchmarks. As summarized in Tab. 1, our methods, U^{2}Flow and U^{2}Flow (+FF), achieve highly competitive results, surpassing all existing unsupervised methods on both KITTI-2015 and Sintel benchmarks. Specifically, on KITTI-2015, U^{2}Flow attains Fl-all=6.13%, significantly outperforming the previous state-of-the-art unsupervised two-frame method UPFlow [29] (9.38%). On KITTI-2012, our method achieves a comparable EPE of 1.4. Furthermore, U^{2}Flow even surpasses approaches that exploit additional information on KITTI-2015, including the multi-frame methods M2Flow [46] (Fl-all=7.37%) and SMURF [41] (Fl-all=6.83%), as well as the semantics-guided methods SemARFlow [54] (Fl-all=8.38%) and UnSAMFlow [55] (Fl-all=7.83%).

The enhanced variant, U^{2}Flow (+FF), further improves performance to an Fl-all=6.00% on KITTI-2015, validating the effectiveness of our bidirectional flow fusion module.

Similarly, on the Sintel benchmark, both U^{2}Flow and U^{2}Flow (+FF) achieve strong results, consistently outperforming prior unsupervised methods across both the clean and final passes.

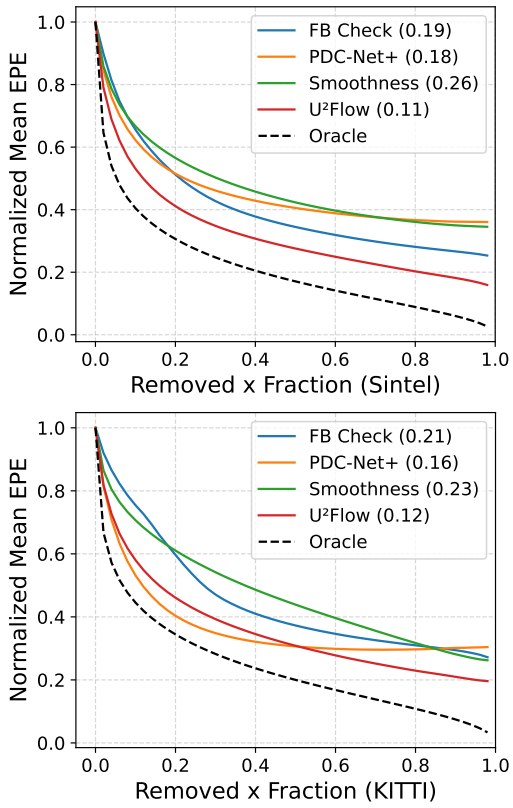


Figure 4. Sparsification curves for uncertainty evaluation. Lower AUSE (shown in parentheses) is better.

4.3. Uncertainty Evaluation

We evaluate the reliability of our uncertainty estimates by measuring their correlation with the ground-truth flow error (EPE) [12, 51]. A high-quality uncertainty prediction should correspond strongly to a large estimation error. To this end, we employ two standard quantitative metrics: the Area Under the Sparsification Error curve (AUSE) [12] and Spearman's Rank Correlation Coefficient (CC) [51].

The AUSE evaluates how well predicted uncertainty serves as a proxy for true error. It is derived from a sparsification process that compares the error curve when removing pixels by predicted uncertainty versus by ground-truth error. Consequently, a lower AUSE indicates a better uncertainty estimation. Spearman's CC directly measures the monotonic relationship between predicted uncertainty and true error, with higher values denoting stronger correlation.

Tab. 2 reports uncertainty estimation on the Sintel and KITTI training sets. U^{2}Flow consistently achieves lower AUSE and higher CC than all baselines, outperforming heuristic methods [1, 32] as well as the model-based PDC-Net+ [48], despite the latter being trained on dense synthetic ground-truth data and fine-tuned on real-world sequences.

The sparsification curves in Fig. 4 provide further insight. They show that U^{2}Flow produces a more robust and

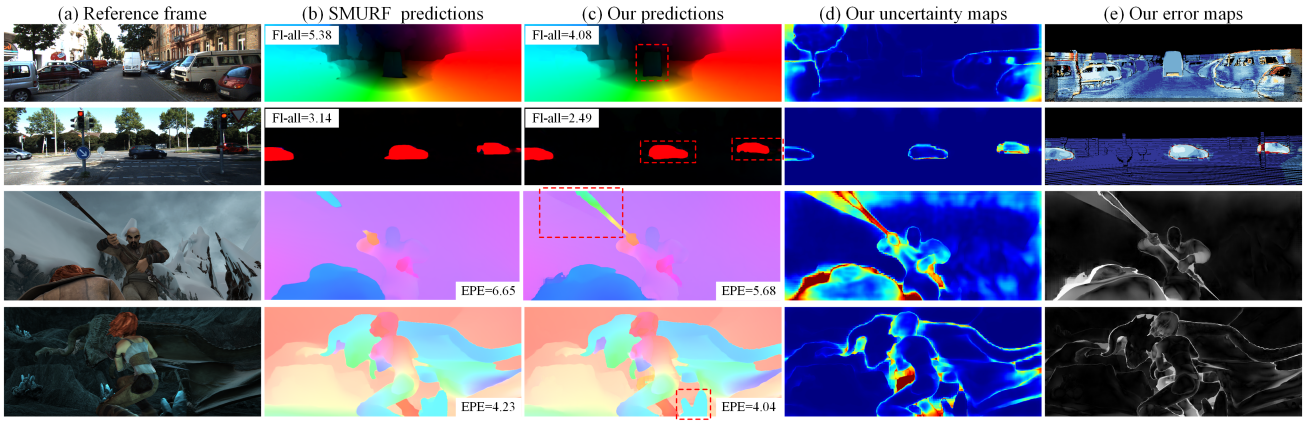


Figure 5. Qualitative results on the KITTI test set (sample frames #5 and #9) and the Sintel (final pass) test set (samples: *ambush_3*, frame 23; *cave_3*, frame 16), compared with SMURF [41]. Additional examples can be found on the official benchmark website.

globally consistent ranking of flow errors on both Sintel and KITTI. While PDC-Net+ slightly outperforms us in the initial portion on KITTI due to real-data fine-tuning, U^{2}Flow's curve remains closer to the oracle across a wider range of removed fractions, resulting in lower final errors and confirming that our predicted uncertainty reliably reflects overall flow quality.

These results demonstrate that U^{2}Flow, guided by augmentation consistency in self-supervised training, effectively captures diverse uncertainty patterns arising from appearance variations and geometric ambiguities.

4.4. Qualitative Results

Qualitative samples from the Sintel and KITTI datasets are presented in Fig. 5. Compared with the state-of-the-art competitor [41], our method generally achieves superior performance, particularly around motion boundaries. Moreover, the predicted uncertainty maps effectively capture and reflect the magnitude of estimation errors.

4.5. Ablation Study

Ablation experiments were performed to assess the contribution of each proposed module, with all models trained under identical conditions except for the ablated components.

Main Components: The effectiveness of our key components is validated through the ablation study summarized in Tab. 3. Introducing only the Uncertainty Estimation (UE) module without a decoupling strategy (row 2) degrades performance across all metrics. This highlights the challenge of joint optimization, where the uncertainty learning objective can interfere with the primary task of flow estimation.

This issue is resolved by incorporating our Decoupling (Dec.) strategy (row 3). By separating the learning objectives for flow and uncertainty, we observe consistent improvements on both Sintel and KITTI. This confirms that decoupling is crucial for preserving the stability of self-supervised learning while still benefiting from the regularizing effect of uncertainty supervision. Further adding the Flow Refinement (FR) module (row 4) continues this trend, demonstrating the benefit of using uncertainty to guide the network's inference process.

UE	Dec.	FR	ℓ_{hg}	FF	Sintel		KITTI 2015	
					Final	Clean	EPE	Fl-all
-	-	-	-	-	2.46	1.56	1.98	6.82
✓	-	-	-	-	2.57	1.65	2.18	7.72
✓	✓	-	-	-	2.41	1.44	1.98	6.87
✓	✓	✓	-	-	2.32	1.42	1.95	6.85
✓	✓	-	✓	-	2.52	1.48	1.87	6.69
✓	✓	✓	◯	-	2.32	1.42	1.83	6.59
✓	✓	✓	◯	✓	**2.29**	**1.36**	**1.74**	**6.30**

Table 3. Ablation study on key components. We evaluate the impact of Uncertainty Estimation (UE), Decoupling of flow and uncertainty learning (Dec.), Flow Refinement (FR), uncertainty-enhanced ℓ_{hg}, and Flow Fusion (FF). The ◯ symbol indicates that ℓ_{hg} is applied only on the KITTI dataset.

The uncertainty-enhanced homography loss (ℓ_{hg}) offers dataset-specific benefits (rows 5 and 6). On KITTI, where scenes frequently involve substantial planar rigid motion, this provides a significant boost in performance. However, on Sintel, which features complex, non-rigid motion, high uncertainty often correlates with dynamic objects where the homography assumption is fragile. Applying this loss can therefore be detrimental. Consequently, we only apply uncertainty-enhanced ℓ_{hg} to the KITTI dataset, as indicated by the ◯ symbol in the table.

Finally, our full model (last row) integrates all components including Flow Fusion (FF). It achieves the best overall performance, reaching 2.29/1.36 on Sintel and 1.74/6.30

Model Variant	Sintel		KITTI 2015	
	Final	Clean	EPE	Fl-all
w/o Refinement	2.41	1.44	1.87	6.69
Refinement w/o Uncertainty	2.42	1.51	1.93	6.81
Refinement w/ Uncertainty	**2.32**	**1.42**	**1.83**	**6.59**

Table 4. Ablation study on the flow refinement module.

Method	Sintel		KITTI 2015		
	Final	Clean	Fl-all	Fl-noc	Fl-occ
w/o Fusion	2.32	1.42	6.59	5.10	16.01
Occ Mask	2.34	1.42	8.17	5.07	27.73
Unc Mask (Ours)	**2.29**	**1.36**	**6.30**	**5.05**	**14.17**

Table 5. Ablation study on the bidirectional flow fusion module.

on KITTI. This result confirms that our complete design can provide uncertainty estimates while delivering optical flow that surpasses the baseline in accuracy.

Flow Refinement Module: We conduct an ablation study to evaluate the proposed uncertainty-aware flow refinement mechanism, which integrates estimated uncertainty to dynamically scale flow features and refine flow estimates (Sec. 3.2). Three variants are defined for comparison: (1) Refinement w/ Uncertainty. (2) w/o Refinement: The model removes the refinement mechanism entirely. The residual is estimated only from the hidden feature, equivalent to standard RAFT-style update: $\Delta\mathbf{F}^{(k)} = \mathcal{C}'_{\text{flow}}(\mathbf{h}^{(k)})$. (3) Refinement w/o Uncertainty: The model retains the refinement structure but excludes the uncertainty branch. The flow residual is refined using only the unscaled flow feature and a dummy zero-map instead of the uncertainty map: $\Delta\mathbf{F}^{(k)} = \mathcal{C}'_{\text{flow}}(\text{concat}(\mathbf{f}^{(k)}, \mathbf{f}^{(k)}, \mathbf{0}))$. This variant isolates the benefit of feature concatenation alone, removing the dynamic influence of uncertainty.

The ablation results are presented in Tab. 4. Performance metrics on the Sintel and KITTI-2015 training sets clearly demonstrate that the refinement architecture, incorporating dynamic scaling based on uncertainty, is crucial for achieving optimal performance.

Effectiveness of Uncertainty-Guided Fusion: The core innovation of our bidirectional flow fusion module is the use of an uncertainty-driven mask, $\mathbf{M}_{\text{fused}}$, to intelligently merge forward and backward flows. To rigorously validate this approach, we compare our full model against two alternatives: one using a traditional occlusion mask for fusion, and another disabling the fusion module entirely.

As shown in Tab. 5, a particularly notable finding is that guiding fusion with a conventional occlusion mask yields results even worse than the baseline without any fusion. This seemingly counterintuitive outcome arises from a fundamental limitation of binary occlusion flags: they fail to differentiate between high- and low-quality flow estimates within the occluded regions. As a result, valid forward flows are often indiscriminately discarded and replaced with backward-warped counterparts of inferior quality, thereby degrading the overall flow field.

In contrast, our uncertainty-based strategy effectively resolves this ambiguity. By assigning fine-grained, per-pixel confidence scores, it enables more informed fusion decisions that preserve reliable flow estimates and improve the handling of high-uncertainty flow in occluded regions. These results demonstrate that the proposed uncertainty mechanism effectively guides downstream flow fusion and validates the reliability of uncertainty estimates.

Method	KITTI→Sintel		Sintel→KITTI	
	Final	Clean	EPE	Fl-all
Ours (w/o Fusion)	5.80	4.67	5.10	17.05
Ours (w/ Fusion)	**5.49**	**4.23**	**4.77**	**16.58**

Table 6. Generalization ability. Training on one dataset and testing directly on the other dataset.

4.6. Generalization Ability

To evaluate the generalization ability of uncertainty estimation, we test whether the uncertainty predicted by a model trained on one domain can still reliably guide the bidirectional flow fusion task when tested directly on another domain without fine-tuning. As shown in Tab. 6, the model with our uncertainty-guided fusion module exhibits significantly stronger domain generalization. This indicates that the reliability signal captured by our uncertainty estimation is fundamental and transferable.

5. Conclusion

We propose U^{2}Flow, a recurrent unsupervised framework for jointly estimating optical flow and per-pixel uncertainty. Our method leverages augmentation consistency and a decoupled learning strategy to achieve stable training of both flow and uncertainty in unsupervised settings. We demonstrate that the learned uncertainty is not merely a byproduct, but a valuable signal that can effectively guide adaptive flow refinement, modulate smoothness constraints, and enable robust bidirectional flow fusion. U^{2}Flow establishes a state-of-the-art for unsupervised optical flow while producing highly reliable uncertainty estimates.

Limitations Our uncertainty supervision relies on the model's predictive variance under a predefined set of augmentations, which may not fully capture all real-world sources of error, such as severe non-Gaussian motion blur or atmospheric distortions. In future work, we plan to enrich the augmentation space using generative models to synthesize more realistic and diverse image degradations.

Acknowledgments

This research was supported by the Guangzhou Basic and Applied Basic Research Foundation under Grant SL2024A04J0183, the Guangxi Key Research and Development Project under Grant GuikeAB25069495, the National Natural Science Foundation of China under Grant 92470202, and the Fund of National Key Laboratory of Multispectral Information Intelligent Processing Technology (No. 202410487201).

References

[1] Rokia Abdein, Wei Li, Chenghao Li, Xiangping Zheng, and Rahul Yadav. Self-supervised uncertainty-guided refinement for robust joint optical flow and depth estimation. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2025. 2, 6

[2] John L. Barron, David J. Fleet, and Steven S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision (IJCV)*, 12(1):43–77, 1994. 2

[3] Andrés Bruhn and Joachim Weickert. A confidence measure for variational optic flow methods. *Computational Imaging and Vision*, page 283, 2006. 2

[4] Daniel J. Butler, Jonas Wulff, Garrett B. Stanley, and Michael J. Black. A naturalistic open source movie for optical flow evaluation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 611–625, 2012. 1, 5

[5] Qiaole Dong and Yanwei Fu. Memflow: Optical flow estimation and prediction with memory. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19068–19078, 2024. 1

[6] 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. 1

[7] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, pages 1050–1059. JMLR.org, 2016. 2

[8] Jochen Gast and Stefan Roth. Lightweight probabilistic deep networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3369–3378, 2018. 2

[9] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research (IJRR)*, 32(11):1231–1237, 2013. 4, 5

[10] Hsin-Ping Huang, Charles Herrmann, Junhwa Hur, Erika Lu, Kyle Sargent, Austin Stone, Ming-Hsuan Yang, and Deqing Sun. Self-supervised autoflow. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11412–11421, 2023. 1

[11] Zhaoyang Huang, Xiaoyu Shi, Chao Zhang, Qiang Wang, Ka Chun Cheung, Hongwei Qin, Jifeng Dai, and Hongsheng Li. Flowformer: A transformer architecture for optical flow. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 668–685, 2022. 1, 5

[12] Eddy Ilg, Özgün Çiçek, Silvio Galesso, Aaron Klein, Osama Makansi, Frank Hutter, and Thomas Brox. Uncertainty estimates and multi-hypotheses networks for optical flow. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 677–693. Springer, 2018. 1, 2, 4, 6

[13] Andrei Iosif and Mihai Negru. Optical flow with semantic guidance and uncertainty estimation for robust video perception. In *Proceedings of the IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*, pages 49–56, 2023. 1, 2, 4

[14] Joel Janai, Fatma Guney, Anurag Ranjan, Michael Black, and Andreas Geiger. Unsupervised learning of multi-frame optical flow with occlusions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 690–706, 2018. 2

[15] Rico Jonschkowski, Austin Stone, Jonathan T Barron, Ariel Gordon, Kurt Konolige, and Anelia Angelova. What matters in unsupervised optical flow. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 557–572, 2020. 1, 2, 5

[16] Jun-Gu Kang, Si-Dong Roh, and Ki-Seok Chung. Flownetu: Accurate uncertainty estimation of optical flow for video object detection. In *Proceedings of the International Conference on Artificial Intelligence and Pattern Recognition (AIPR)*, pages 36–41, 2022. 1, 2, 4

[17] Hannah Halin Kim, Shuzhi Yu, Shuai Yuan, and Carlo Tomasi. Cross-attention transformer for video interpolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 320–337, 2022. 1

[18] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[19] Claudia Kondermann, Rudolf Mester, and Christoph Garbe. A statistical confidence measure for optical flows. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 290–301. Springer, 2008. 2

[20] Jan Kybic and Claudia Nieuwenhuis. Bootstrap optical flow and uncertainty measure. *Computer Vision and Image Understanding (CVIU)*, 115(10):1449–1462, 2011. 2

[21] Guillaume Le Moing, Jean Ponce, and Cordelia Schmid. Dense optical tracking: Connecting the dots. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19187–19197, 2024. 1

[22] Wenbin Lin, Chengwei Zheng, Jun-Hai Yong, and Feng Xu. Occlusionfusion: Occlusion-aware motion estimation for real-time dynamic 3d reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1736–1745, 2022. 1

[23] Jiuming Liu, Guangming Wang, Weicai Ye, Chaokang Jiang, Jinru Han, Zhe Liu, Guofeng Zhang, Dalong Du, and Hesheng Wang. Difflow3d: Toward robust uncertainty-aware

scene flow estimation with iterative diffusion-based refinement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15109–15119, 2024. 2

[24] Liang Liu, Jiangning Zhang, Ruifei He, Yong Liu, Yabiao Wang, Ying Tai, Donghao Luo, Chengjie Wang, Jilin Li, and Feiyue Huang. 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 (CVPR)*, pages 6488–6497, 2020. 1, 2, 3, 4, 5, 6

[25] Pengpeng Liu, Irwin King, Michael R. Lyu, and Jia Xu. Ddflow: Learning optical flow with unlabeled data distillation. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 8770–8777, 2019. 2

[26] Pengpeng Liu, Michael Lyu, Irwin King, and Jia Xu. Selflow: Self-supervised learning of optical flow. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4571–4580, 2019. 1, 2, 5

[27] Pengpeng Liu, Irwin King, Michael R Lyu, and Jia Xu. Flow2stereo: Effective self-supervised learning of optical flow and stereo matching. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6648–6657, 2020. 1

[28] 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 (CVPR)*, pages 19167–19176, 2024. 1, 5

[29] Kunming Luo, Chuan Wang, Shuaicheng Liu, Haoqiang Fan, Jue Wang, and Jian Sun. Upflow: Upsampling pyramid for unsupervised optical flow learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1045–1054, 2021. 1, 2, 5, 6

[30] Oisin Mac Aodha, Ahmad Humayun, Marc Pollefeys, and Gabriel J Brostow. Learning a confidence measure for optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 35(5):1107–1120, 2012. 2

[31] Rémi Marsal, Florian Chabot, Angélique Loesch, and Hichem Sahbi. Brightflow: Brightness-change-aware unsupervised learning of optical flow. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 2061–2070, 2023. 1, 2

[32] Simon Meister, Junhwa Hur, and Stefan Roth. Unflow: Unsupervised learning of optical flow with a bidirectional census loss. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2018. 1, 2, 3, 5, 6

[33] Moritz Menze and Andreas Geiger. Object scene flow for autonomous vehicles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3061–3070, 2015. 4, 5

[34] Matteo Poggi, Filippo Aleotti, Fabio Tosi, and Stefano Mattoccia. On the uncertainty of self-supervised monocular depth estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3227–3237, 2020. 1

[35] Zhe Ren, Junchi Yan, Bingbing Ni, Bin Liu, Xiaokang Yang, and Hongyuan Zha. Unsupervised deep learning for optical flow estimation. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2017. 2

[36] Shihao Shen, Louis Kerofsky, and Senthil Yogamani. Optical flow for autonomous driving: Applications, challenges and improvements. *arXiv preprint arXiv:2301.04422*, 2023. 1

[37] Yichen Shen, Zhilu Zhang, Mert R. Sabuncu, and Lin Sun. Real-time uncertainty estimation in computer vision via uncertainty-aware distribution distillation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 707–716, 2021. 1, 2, 4

[38] Xiaoyu Shi, Zhaoyang Huang, Weikang Bian, Dasong Li, Manyuan Zhang, Ka Chun Cheung, Simon See, Hongwei Qin, Jifeng Dai, and Hongsheng Li. Videoflow: Exploiting temporal cues for multi-frame optical flow estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12435–12446, 2023. 1, 5

[39] Xiaoyu Shi, Zhaoyang Huang, Dasong Li, Manyuan Zhang, Ka Chun Cheung, Simon See, Hongwei Qin, Jifeng Dai, and Hongsheng Li. Flowformer++: Masked cost volume autoencoding for pretraining optical flow estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1599–1610, 2023. 1

[40] Leslie N Smith and Nicholay Topin. Super-convergence: Very fast training of neural networks using large learning rates. In *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, pages 369–386, 2019. 5

[41] Austin Stone, Daniel Maurer, Alper Ayvaci, Anelia Angelova, and Rico Jonschkowski. 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)*, pages 3887–3896, 2021. 2, 5, 6, 7

[42] 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/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8934–8943, 2018. 1

[43] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. Models matter, so does training: An empirical study of cnns for optical flow estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 42(6):1408–1423, 2019. 1, 5

[44] Deqing Sun, Daniel Vlasic, Charles Herrmann, Varun Jampani, Michael Krainin, Huiwen Chang, Ramin Zabih, William T Freeman, and Ce Liu. Autoflow: Learning a better training set for optical flow. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10093–10102, 2021. 1

[45] Deqing Sun, Charles Herrmann, Fitsum Reda, Michael Rubinstein, David J. Fleet, and William T. Freeman. Disentangling architecture and training for optical flow. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 165–182, 2022. 1

[46] Xunpei Sun, Gang Chen, and Zuoxun Hou. M2flow: A motion information fusion framework for enhanced unsupervised optical flow estimation in autonomous driving. In *Pro-*

ceedings of the AAAI Conference on Artificial Intelligence (AAAI), pages 7140–7148, 2025. 1, 2, 5, 6

[47] Zachary Teed and Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 402–419, 2020. 1, 2, 3, 5

[48] Prune Truong, Martin Danelljan, Radu Timofte, and Luc Van Gool. Pdc-net+: Enhanced probabilistic dense correspondence network. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 45(8):10247–10266, 2023. 2, 6

[49] Fangjinhua Wang, Silvano Galliani, Christoph Vogel, and Marc Pollefeys. Itermvs: Iterative probability estimation for efficient multi-view stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8606–8615, 2022. 2

[50] Yang Wang, Yi Yang, Zhenheng Yang, Liang Zhao, Peng Wang, and Wei Xu. Occlusion aware unsupervised learning of optical flow. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4884–4893, 2018. 1, 2

[51] Anne S Wannenwetsch, Margret Keuper, and Stefan Roth. Probflow: Joint optical flow and uncertainty estimation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 1173–1182, 2017. 1, 2, 6

[52] Zhichao Yin, Trevor Darrell, and Fisher Yu. Hierarchical discrete distribution decomposition for match density estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6037–6046, 2019. 2

[53] Shuai Yuan, Xian Sun, Hannah Kim, Shuzhi Yu, and Carlo Tomasi. Optical flow training under limited label budget via active learning. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 410–427, 2022. 1

[54] Shuai Yuan, Shuzhi Yu, Hannah Kim, and Carlo Tomasi. Semarflow: Injecting semantics into unsupervised optical flow estimation for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9566–9577, 2023. 2, 5, 6

[55] Shuai Yuan, Lei Luo, Zhuo Hui, Can Pu, Xiaoyu Xiang, Rakesh Ranjan, and Denis Demandolx. Unsamflow: Unsupervised optical flow guided by segment anything model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19027–19037, 2024. 2, 3, 4, 5, 6

[56] Xinyu Zhang and Abdeslam Boularias. Optical flow boosts unsupervised localization and segmentation. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7635–7642, 2023. 1

[57] Yushan Zhang, Bastian Wandt, Maria Magnusson, and Michael Felsberg. Diffsf: Diffusion models for scene flow estimation. *Advances in Neural Information Processing Systems (NeurIPS)*, 37:11227–11247, 2024. 2

[58] Zhihang Zhong, Gurunandan Krishnan, Xiao Sun, Yu Qiao, Sizhuo Ma, and Jian Wang. Clearer frames, anytime: Resolving velocity ambiguity in video frame interpolation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 346–363, 2024. 1

[59] Kaichen Zhou, Jia-Wang Bian, Jian-Qing Zheng, Jiaxing Zhong, Qian Xie, Niki Trigoni, and Andrew Markham. Manydepth2: Motion-aware self-supervised monocular depth estimation in dynamic scenes. *IEEE Robotics and Automation Letters*, 2025. 1

[60] Mo Zhou, Jingwei Wang, Xinyu Zhang, Dylan Campbell, Kaixuan Wang, Li Yuan, and Xiao Lin. Probdiffflow: An efficient learning-free framework for probabilistic single-image optical flow estimation. *arXiv preprint arXiv:2503.12348*, 2025. 2