Optical Flow Training Under Limited Label Budget via Active Learning

Shuai Yuan(✉), Xian Sun, Hannah Kim, Shuzhi Yu, and Carlo Tomasi

Duke University, Durham, NC 27708, USA
{shuai,hannah,shuzhiyu,tomasi}@cs.duke.edu, xian.sun@duke.edu

Abstract. Supervised training of optical flow predictors generally yields better accuracy than unsupervised training. However, the improved performance comes at an often high annotation cost. Semi-supervised training trades off accuracy against annotation cost. We use a simple yet effective semi-supervised training method to show that even a small fraction of labels can improve flow accuracy by a significant margin over unsupervised training. In addition, we propose active learning methods based on simple heuristics to further reduce the number of labels required to achieve the same target accuracy. Our experiments on both synthetic and real optical flow datasets show that our semi-supervised networks generally need around 50% of the labels to achieve close to full-label accuracy, and only around 20% with active learning on Sintel. We also analyze and show insights on the factors that may influence active learning performance. Code is available at https://github.com/duke-vision/optical-flow-active-learning-release.

Keywords: Optical flow · Active learning · Label efficiency

1 Introduction

The estimation of optical flow is a very important but challenging task in computer vision with broad applications including video understanding [7], video editing [9], object tracking [1], and autonomous driving [34].

Inspired by the successes of deep CNNs in various computer vision tasks [12, 23], much recent work has modeled optical flow estimation in the framework of supervised learning, and has proposed several networks of increasingly high performance on benchmark datasets [5,14,15,38,47,49,60]. Ground-truth labels provide a strong supervision signal when training these networks. However, ground-truth optical flow annotations are especially hard and expensive to obtain. Thus, many methods use synthetic data in training, since ground-truth

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-20047-2_24.

S. Avidan et al. (Eds.): ECCV 2022, LNCS 13682, pp. 410–427, 2022.
https://doi.org/10.1007/978-3-031-20047-2_24

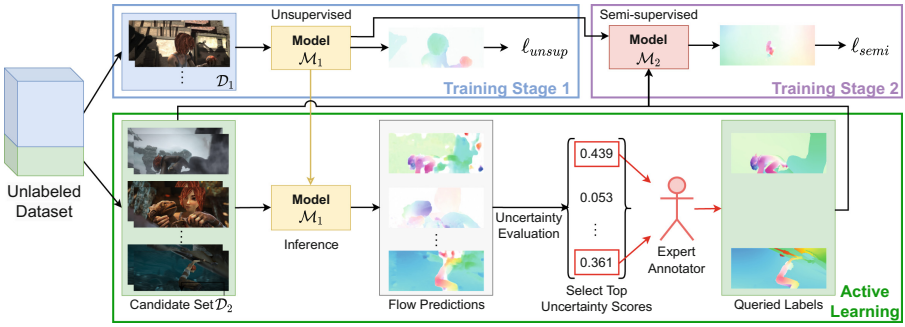


Fig. 1. Overview of our active learning framework for the semi-supervised training.

labels can be generated as part of data synthesis. Nevertheless, it is still an open question whether synthetic data are an adequate proxy for real data.

Another way to circumvent label scarcity is unsupervised training, which does not require any labels at all. Instead, it relies on unsupervised loss measures that enforce exact or approximate constraints that correct outputs should satisfy. Common losses used in unsupervised optical flow estimation are the photometric loss, which penalizes large color differences between corresponding points, and the smoothness loss, which penalizes abrupt spatial changes in the flow field [17, 18, 28, 29, 33, 40]. While unsupervised methods allow training on large datasets from the application domain, their performance is still far from ideal because the assumed constraints do not always hold. For instance, the photometric loss works poorly with non-Lambertian surfaces or in occlusion regions [52], while the smoothness loss fails near motion discontinuities [21].

Semi-supervised training can be a way to combine the advantages of both supervised and unsupervised training for optical flow models. The idea is simple, and amounts to training the network with a mix of labeled and unlabeled data. This is possible because we can charge different losses (supervised or unsupervised) to different samples depending on whether they are labeled or not.

The trade-off between performance and labeling cost is of interest in real practice, since it describes the marginal benefit that can be accrued at the price of a unit of labeling effort. However, little work has focused on the semi-supervised training of optical flow. Existing methods have tried to improve flow estimates given an available, partially labeled dataset [24, 45, 57]. Other work uses semi-supervised training to address specific problem conditions, *e.g.,* foggy scenes [56].

In contrast, we are particularly interested in label efficiency, that is, in the performance improvement gained as the fraction of labeled samples increases from 0 ("unsupervised") to 1 ("supervised"). Specifically, we use a simple yet effective semi-supervised algorithm and show that the model error drops significantly as soon as a small fraction of the samples are labeled. This suggests that even a modest labeling budget can lead to a significant performance boost.

Given a specific labeling budget, an important related question is how to determine which part of the dataset to label. A simple method is random sampling, but it is possible to do better. Specifically, we propose and evaluate

criteria that suggest whose labels bring larger benefits in training. This brings us to the concept of active learning.

Active Learning (AL) has been shown to be effective in reducing annotation costs while maintaining good performance in many vision tasks including image classification [2,27], object detection [4,42], semantic segmentation [31,44], and instance segmentation [51]. The general idea is to allow the training algorithm to select valuable unlabeled samples for which to query labels for further training. This selection is especially important for optical flow estimation, since generating labels for additional samples incurs high costs in terms of computation, curation, and sometimes even hand annotations.

While annotating individual flow vectors by hand is effectively impossible in practice, annotation can be and often is done by hand at a higher level and, even so, is costly. For instance, in KITTI 2015 [34], correspondences between points on CAD models of moving cars are annotated by hand so that dense optical flow can be inferred for these cars. In addition, nonrigid objects such as pedestrians or bicyclists are manually masked out, and so are errors in the flow and disparity masks inferred from LiDAR and GPS/IMU measurements and from stereo depth estimation. This is still manual annotation and curation, painstaking and expensive. Some amount of curation, at the very least, is necessary for most high-quality training sets with real imagery, and the methods we propose aim to reduce the need for this type of work, and to make the products of whatever manual work is left more effective. To the best of our knowledge, we are the first to study active learning as a way to moderate the high annotation costs for optical flow estimation.

As illustrated in Fig. 1, our training pipeline (top part of the diagram) includes an unsupervised first stage and a semi-supervised second stage. We split our unlabeled dataset to two sets, one ($\mathcal{D}_1$) used to pre-train an unsupervised model $\mathcal{M}_1$ and the other ($\mathcal{D}_2$) used as the *candidate* set, from which samples are selected to query labels from expert annotators. After training model $\mathcal{M}_1$ on $\mathcal{D}_1$ in Stage 1, we estimate flow for all the samples in $\mathcal{D}_2$ and score each of them based on our active learning criteria. We query for labels for top-scoring samples and add these to $\mathcal{D}_2$ for further semi-supervised training in Stage 2. In this paper, we show that using active learning to query labels can help further reduce the number of labels required to achieve a given performance target in semi-supervised training.

In summary, our contributions are as follows.

- We show on several synthetic and real-life datasets that the performance from unsupervised training of optical flow estimators can be improved significantly as soon as a relatively small fraction of labels are added for semi-supervised training.
- To the best of our knowledge, we are the first to explore active learning as a way to save annotation cost for optical flow estimation, and our novel pipeline can be used directly in real practice.
- We set up the new problem of semi-supervised training of optical flow under certain label ratio constraints. We anticipate follow-up research to propose better methods for this problem.

2 Related Work

Supervised Optical Flow. Supervised methods use deep networks to learn the mapping from image pairs to the corresponding optical flow by minimizing the supervised loss, namely, some distance measure between computed and true flow. FlowNet [5] used a multi-scale encoder-decoder structure with skip connections between same-scale layers. Following this framework, many networks have been proposed to decrease both model size and error. Traditional ideas or heuristics have been introduced into the network, including image pyramid in SPyNet [38], feature pyramid, warping, and cost volume in PWC-Net [47] and LiteFlowNet [14]. Iterative decoder modules have also been explored as a way to reduce model size while retaining accuracy in IRR-PWC [15] and RAFT [49]. The latter built the network based on full-pair correlations and has led to many follow-up models that have achieved the state-of-the-art performance [60].

Unsupervised Optical Flow. Recent research has focused on the unsupervised learning of optical flow as a compromise between label availability and model performance. Initial work on this topic proposed to train FlowNet-like networks using surrogate loss terms, namely photometric loss and smoothness loss [18,40]. As found by many papers, flow at occlusion region is especially challenging for unsupervised networks [52]. Thus, much research focused on solving the occlusion problem via occlusion masks [52], bi-directional consistency [33], multi-frame consistency [17,41], and self-supervised teacher-student models [29,59]. ARFlow [28] integrated a second forward pass using transformed inputs for augmentation and has achieved the state-of-the-art unsupervised performance. Multi-frame unsupervised models have also been investigated [17,46].

Semi-supervised Training in Vision. Semi-supervised training targets applications where partial labels are available. Early approaches in image classification [11,25,35,43] utilize label propagation with regularization and augmentation based on the belief that nearby data points tend to have similar class labels. A more recent class of methods train on unlabeled samples with pseudo-labels [26,55] predicted by a supervised trained network trained with labeled samples. Similar teacher-student models have also been explored [30,48].

Although widely explored in many other vision tasks, there is little work on semi-supervised optical flow. Some early work utilized semi-supervised learning to achieve comparable flow accuracy to the supervised methods [24,24,45,57]. Others applied semi-supervised methods to tackle specific cases of optical flow, such as dense foggy scenes [56] and ultrasound elastography [50]. In contrast, we focus on label efficiency for optical flow estimation: Instead of proposing semi-supervised networks that focus on improving benchmark performances by adding external unlabeled data, we are more focused on the trade-off between performance and label ratio given a fixed dataset.

Active Learning in Vision. Active Learning (AL) aims to maximize model performance with the least amount of labeled data by keeping a human in the training loop. The general idea is to make the model actively select the most

valuable unlabeled samples and query the human for labels which are used in the next stage of training. There are two main categories, namely, uncertainty-based (select samples based on some pre-defined uncertainty metric) [6,8,13,20], and distribution-based (query representative samples of sufficient diversity) [37,54].

Active learning has achieved extensive success in various fields in computer vision, including image classification [2,27], object detection [4,42], semantic segmentation [31,44], and instance segmentation [51]. However, the concept has received little attention in optical flow estimation where acquiring labels is especially difficult. To the best of our knowledge, we are the first to apply active learning to optical flow estimation to reduce annotation cost.

3 Method

As we are among the first to explore active learning as a way to tackle the high annotation costs in optical flow training, we start from simple yet effective methods to implement our ideas. This section describes our semi-supervised training method (Sect. 3.1), active learning heuristics (Sect. 3.2), and network structure and loss functions (Sect. 3.3).

3.1 Semi-supervised Training

Given a partially labeled data set, we implement the semi-supervised training by charging a supervised loss to the labeled samples and an unsupervised loss to the unlabeled ones. Specifically, the semi-supervised loss for each sample $\boldsymbol{x}$ is

$$\ell_{\text{semi}}(\boldsymbol{x}) = \begin{cases} \ell_{\text{unsup}}(\boldsymbol{x}), \text{ if } \boldsymbol{x} \text{ is unlabeled,} \\ \alpha \ell_{\text{sup}}(\boldsymbol{x}), \text{ otherwise} \end{cases} \tag{1}$$

where $\alpha > 0$ is a balancing weight. We do not include the unsupervised loss for labeled samples (although in principle this is also an option) to avoid any conflict between the two losses, especially on occlusion and motion boundary regions.

Thus, the final loss of the data set $\mathcal{D} = \mathcal{D}^u \cup \mathcal{D}^l$ is

$$\mathcal{L}_{\text{semi}} = \sum_{\boldsymbol{x} \in \mathcal{D}} \ell_{\text{semi}}(\boldsymbol{x}) = \sum_{\boldsymbol{x} \in \mathcal{D}^u} \ell_{\text{unsup}}(\boldsymbol{x}) + \alpha \sum_{\boldsymbol{x} \in \mathcal{D}^l} \ell_{\text{sup}}(\boldsymbol{x}), \tag{2}$$

where $\mathcal{D}^u$ and $\mathcal{D}^l$ are the unlabeled and labeled sets. We define the *label ratio* as $r = |\mathcal{D}^l|/|\mathcal{D}|$. During training, we randomly shuffle the training set $\mathcal{D}$, so that each batch of data has a mix of labeled and unlabeled samples.

3.2 Active Learning Heuristics

Figure 1 shows a general overview of our active learning framework. After pre-training our model on unlabeled data (Stage 1), we invoke an active learning algorithm to determine samples to be labeled for further training. Specifically, we first use the pre-trained model to infer flow on the samples of another disjoint

unlabeled data set (the *candidate* set) and select a fraction of the samples to be labeled, based on some criterion. After obtaining those labels, we continue to train the model on the partially labeled candidate set using the semi-supervised loss (Stage 2). Note that in this second stage, we do not include the unlabeled data used in pre-training (see ablation study in Sect. 4.5). By allowing the model to actively select samples to query labels, we expect the model to achieve the best possible performance under a fixed ratio of label queries (the "label budget").

So, what criteria should be used for selecting samples to be labeled? Many so-called uncertainty-based methods for active learning algorithms for image classification or segmentation use the soft-max scores to compute how confident the model is about a particular output. However, optical flow estimation is a regression problem, not a classification problem, so soft-max scores are typically not available, and would be in any case difficult to calibrate.

Instead, we select samples for labeling based on heuristics specific to the optical flow problem. For example, the photometric loss is low for good predictions. In addition, unsupervised flow estimation performs poorly at occlusion regions and motion discontinuities. These considerations suggest the following heuristic metrics to flag points for which unsupervised estimates of flow are poor:

- *Photo loss*: the photometric loss used in training.
- *Occ ratio*: the ratio of occlusion pixels in the frame, with occlusion estimated by consistency check of forward and backward flows [33].
- *Flow grad norm*: the magnitude of gradients of the estimated flow field as in [16] averaged across the frame, used to indicate the presence of motion boundaries.

We experiment with three active learning methods, each using one of the metrics above. When querying labels for a given label ratio r, we first compute the metric for each sample in the candidate set, and then sort and pick the samples with largest uncertainties as our queries.

3.3 Network Structure and Loss Functions

Network Structure. We adopt the unsupervised state-of-the-art, ARFlow [28], as our base network, which is basically a lightweight variant of PWC-Net [47]. PWC-Net-based structures have been shown to be successful in both supervised and unsupervised settings, so it is a good fit for our hybrid semi-supervised training. We do not choose RAFT because it has been mostly proven to work well in the supervised setting, while our setting (Sect. 4.4) is much closer to the unsupervised one (see appendix for details).

Each sample is a triple $\boldsymbol{x} = (I_1, I_2, U_{12})$ where $I_1, I_2 \in \mathbb{R}^{h \times w \times 3}$ are the two input frames and U_{12} is the true optical flow (set as "None" for unlabeled samples). The network estimates a multi-scale forward flow field $f(I_1, I_2) = \{\hat{U}_{12}^{(2)}, \hat{U}_{12}^{(3)}, \cdots, \hat{U}_{12}^{(6)}\}$, where the output $\hat{U}_{12}^{(l)}$ at scale l has dimension $\frac{h}{2^l} \times \frac{w}{2^l} \times 2$. The finest estimated scale is $\hat{U}_{12}^{(2)}$, which is up-sampled to yield the final output.

Unsupervised Loss. For unsupervised loss $\ell_{\text{unsup}}(\boldsymbol{x})$ we follow ARFlow [28], which includes a photometric loss $\ell_{\text{ph}}(\boldsymbol{x})$, a smoothness loss $\ell_{\text{sm}}(\boldsymbol{x})$, and an augmentation loss $\ell_{\text{aug}}(\boldsymbol{x})$:

$$\ell_{\text{unsup}}(\boldsymbol{x}) = \ell_{\text{ph}}(\boldsymbol{x}) + \lambda_{\text{sm}}\ell_{\text{sm}}(\boldsymbol{x}) + \lambda_{\text{aug}}\ell_{\text{aug}}(\boldsymbol{x}). \tag{3}$$

Specifically, given the sample $\boldsymbol{x}$, we first estimate both forward and backward flow, $\hat{U}_{12}^{(l)}$ and $\hat{U}_{21}^{(l)}$, and then apply forward-backward consistency check [33] to estimate their corresponding occlusion masks, $\hat{O}_{12}^{(l)}$ and $\hat{O}_{21}^{(l)}$.

To compute the photometric loss, we first warp the frames by $\hat{I}_1^{(l)}(\boldsymbol{p}) = I_2^{(l)}(\boldsymbol{p} + \hat{U}_{12}^{(l)}(\boldsymbol{p}))$, where $I_2^{(l)}$ is I_2 down-sampled to the l-th scale and $\boldsymbol{p}$ denotes pixel coordinates at that scale. The occlusion-aware photometric loss at each scale can be then defined as

$$\ell_{\text{ph}}^{(l)}(\boldsymbol{x}) = \sum_{i=1}^{3} c_i \ \rho_i(\hat{I}_1^{(l)}, I_1^{(l)}, \hat{O}_{12}^{(l)}) \tag{4}$$

where ρ_1, ρ_2, ρ_3 are three distance measures with the estimated occlusion region filtered out in computation. As proposed in [28], these three measures are the L_1-norm, structural similarity (SSIM) [53], and the ternary census loss [33], respectively, weighted by c_i.

The edge-aware smoothness loss of each scale l is computed using the second-order derivatives:

$$\ell_{\text{sm}}^{(l)}(\boldsymbol{x}) = \frac{1}{2|\Omega^{(l)}|} \sum_{z\in\{x,y\}} \sum_{\boldsymbol{p}\in\Omega^{(l)}} \left\| \frac{\partial^2 \hat{U}_{12}^{(l)}(\boldsymbol{p})}{\partial z^2} \right\|_1 e^{-\delta \left\| \frac{\partial I_1(\boldsymbol{p})}{\partial z} \right\|_1}, \tag{5}$$

where $\delta = 10$ is a scaling parameter, and $\Omega^{(l)}$ denotes the set of pixel coordinates on the l-th scale.

We combine the losses of each scale linearly using weights $w_{\text{ph}}^{(l)}$ and $w_{\text{sm}}^{(l)}$ by

$$\ell_{\text{ph}}(\boldsymbol{x}) = \sum_{l=2}^{6} w_{\text{ph}}^{(l)} \ell_{\text{ph}}^{(l)}(\boldsymbol{x}), \quad \ell_{\text{sm}}(\boldsymbol{x}) = \sum_{l=2}^{6} w_{\text{sm}}^{(l)} \ell_{\text{sm}}^{(l)}(\boldsymbol{x}). \tag{6}$$

We also include the photometric and smoothness loss for the backward temporal direction, which is not shown here for conciseness.

After the first forward pass of the network, ARFlow also conducts an additional forward pass on input images transformed with random spatial, appearance, and occlusion transformations to mimic online augmentation. The augmentation loss $\ell_{\text{aug}}(\boldsymbol{x})$ is then computed based on the consistency between outputs before and after the transformation. See [28] for details.

Supervised Loss. For supervised loss $\ell_{\text{sup}}(\boldsymbol{x})$, we apply the multi-scale robust L_1-norm

$$\ell_{\text{sup}}(\boldsymbol{x}) = \sum_{l=2}^{6} \frac{w_{\text{sup}}^{(l)}}{|\Omega^{(l)}|} \sum_{\boldsymbol{p}\in\Omega^{(l)}} (\|\hat{U}_{12}^{(l)}(\boldsymbol{p}) - U_{12}^{(l)}(\boldsymbol{p})\|_1 + \epsilon)^q, \tag{7}$$

where $U_{12}^{(l)}$ is the down-sampled true flow to the l-th scale. A small ϵ and $q < 1$ is included to penalize less on outliers. We set $\epsilon = 0.01$ and $q = 0.4$ as in [47].

Semi-supervised Loss. The semi-supervised loss is computed by Eq. (1).

4 Experimental Results

4.1 Datasets

As most optical flow methods, we train and evaluate our method on FlyingChairs [5], FlyingThings3D [32], Sintel [3], and KITTI [10,34] datasets. Apart from the labeled datasets, raw Sintel and KITTI frames with no labels are also available and often used in recent unsupervised work [28,29,39,58]. As common practice, we have excluded the labeled samples from the raw Sintel and KITTI datasets.

In our experiments, we also split our own train and validation set on Sintel and KITTI. We split Sintel clean and final passes by scenes to 1,082 training samples and 1,000 validation samples. For KITTI, we put the first 150 samples in each of 2015 and 2012 set as our training set, yielding 300 training samples and 94 validation samples. A summary of our data splits is in the appendix.

4.2 Implementation Details

We implement the model in PyTorch [36], and all experiments share the same hyper-parameters as follows. Training uses the Adam optimizer [22] with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and batch size 8. The balancing weight α in Eq. (1) is set as 1. The weights of each unsupervised loss term in Eq. (3) are $\lambda_{\text{sm}} = 50$ for Sintel and $\lambda_{\text{sm}} = 75$ otherwise; and $\lambda_{\text{aug}} = 0.2$ unless otherwise stated. The weights of different distance measures in Eq. (4) are set as $(c_1, c_2, c_3) = (0.15, 0.85, 0)$ in the first 50k iterations and $(c_1, c_2, c_3) = (0, 0, 1)$ in the rest as in ARFlow [28].

The supervised weights $w_{\text{sup}}^{(l)}$ for scales $l = 2, 3, \cdots, 6$ in Eq. (7) are 0.32, 0.08, 0.02, 0.01, 0.005 as in PWC-Net [47]. The photometric weights $w_{\text{ph}}^{(l)}$ in Eq. (6) are 1, 1, 1, 1, 0, and the smoothness weights $w_{\text{sm}}^{(l)}$ in Eq. (6) are 1, 0, 0, 0, 0.

For data augmentation, we include random cropping, random rescaling, horizontal flipping, and appearance transformations (brightness, contrast, saturation, hue, Gaussian blur). Please refer to the appendix for more details.

4.3 Semi-supervised Training Settings

The goal of this first experiment is to see how the validation error changes as we gradually increase the label ratio r from 0 (unsupervised) to 1 (supervised). We are specifically interested in the changing error rate, which reflects the marginal gain of a unit of labeling effort.

We ensure that all experiments on the same dataset have exactly the same setting except the label ratio r for fair comparison. For each experiment, the

labeled set is sampled uniformly. We experiment on all four datasets independently using label ratio $r \in \{0, 0.05, 0.1, 0.2, 0.4, 0.6, 0.8, 1\}$ with settings below.

FlyingChairs and FlyingThings3D. As a simple toy experiment, we split the labeled and unlabeled sets randomly and train using the semi-supervised loss. We train for 1,000k iterations with a fixed learning rate $\eta = 0.0001$.

Sintel. Unlike the two large datasets above, Sintel only has ground-truth labels for 2,082 clean and final samples, which is too small to train a flow model effectively on its own. Thus, the single-stage schedule above may not apply well.

Instead, as is common practice in many unsupervised methods, we first pre-train the network using the large Sintel raw movie set in an unsupervised way. Subsequently, as the second stage, we apply semi-supervised training with different label ratios on our training split of clean and final samples. Note that we compute the label ratio r as the ratio of labeled samples only in our second-stage train split, which does not include the unlabeled raw data samples in the first stage. This is because the label ratio would otherwise become too small (thus less informative) since the number of raw data far exceeds clean and final data.

We train the first stage using learning rate $\eta = 0.0001$ for 500k iterations, while the second stage starts with $\eta = 0.0001$, which is cut by half at 400, 600, and 800 epochs, and ends at 1,000 epochs. Following ARFlow [28], we turn off the augmentation loss by assigning $\lambda_{\text{aug}} = 0$ in the first stage.

KITTI. We apply a similar two-stage schedule to KITTI. We first pre-train the network using KITTI raw sequences with unsupervised loss. Subsequently, we assign labels to our train split of the KITTI 2015/2012 set with a given label ratio by random sampling and then run the semi-supervised training. The learning rate schedule is the same as that for Sintel above.

4.4 Active Learning Settings

The second part of experiments is on active learning, where we show that allowing the model to select which samples to label can help reduce the error.

We mainly experiment on Sintel and KITTI since they are close to real data. Since active learning is a multi-stage process (which needs a pre-trained model to query labels for the next stages), it fits well with the two-stage semi-supervised settings described in Sect. 4.3. Thus, we use those settings with labels queried totally at random as our baseline. In comparison, we show that using the three active learning heuristics described in Sect. 3.2 to query labels can yield better results than random sampling. We try small label ratios $r \in \{0.05, 0.1, 0.2\}$ since the semi-supervised training performance starts to saturate at larger label ratios.

4.5 Main Results

Semi-supervised Training. We first experiment with the semi-supervised training with different label ratios across four commonly used flow datasets. As shown in Fig. 2, the model validation error drops significantly at low label

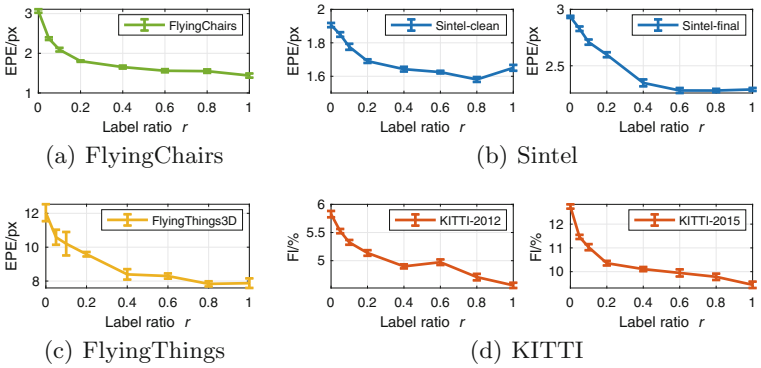


Fig. 2. Model validation errors of the semi-supervised training with different label ratios. 'EPE': End-Point Error, 'Fl': Flow error percentage.

ratios and tends to saturate once an adequate amount of labels are used. This supports our hypothesis that even a few labels can help improve performance significantly.

Another observation is that the errors for FlyingChairs, FlyingThings3D, and Sintel saturate at around 50% labeling, whereas KITTI keeps improving slowly at high label ratios. One explanation for this discrepancy may involve the amount of repetitive information in the dataset: Sintel consists of video sequences with 20–50 frames that are very similar to each other, while KITTI consists of individually-selected frame pairs independent from the other pairs.

Active Learning. Our active learning results are shown in Fig. 3. We compare the validation errors for our three active learning criteria against the baseline setting, in which the labeled samples are selected randomly. To better illustrate the scale of the differences, we add two horizontal lines to indicate totally unsupervised and supervised errors as the "upper" and "lower" bound, respectively.

The Sintel results (Fig. 3a) show that all our three active learning algorithms can improve the baseline errors by large margins. Notably, our active learning algorithms can achieve close to supervised performance with only 20% labeling. This number is around 50% without active learning.

The KITTI results (Fig. 3b) show slight improvements with active learning. Among our three algorithms, "occ ratio" works consistently better than random sampling, especially at a very small label ratio $r = 0.05$. We discuss the reason why our active learning methods help less on KITTI at the end of this chapter.

Among our three active learning heuristics, "occ ratio" has the best performance overall and is therefore selected as our final criterion. Note that the occlusion ratio is computed via a forward-backward consistency check, so it captures not only real occlusions but also inconsistent flow estimates.

Benchmark Testing. We also show results on the official benchmark test sets. Qualitative examples are also included in the appendix. As is shown in Table 1,

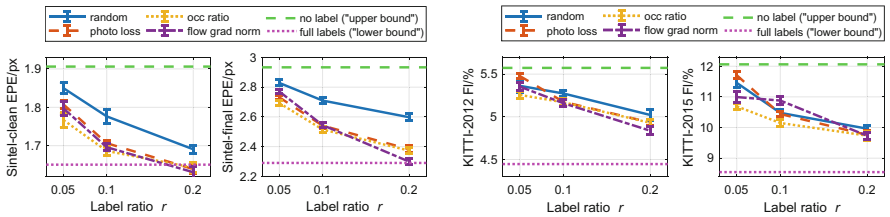


Fig. 3. Validation errors of different active learning algorithms compared with random sampling (baseline); pseudo error bars obtained by taking the standard deviations in the last 50 epoches.

compared with the backbone ARFlow [28] and two other top unsupervised estimators [19,29], our Sintel test EPEs improve significantly even when we utilize a very small fraction (5–20%) of labels in training. This holds true for both clean and final passes, as well as occluded and non-occluded pixels. To indicate the scale of improvements, our semi-supervised results are even comparable to the supervised IRR-PWC [15], which has a similar PWC-Net-based structure, even if we only use 20% of the Sintel labels. We also include the state-of-the-art RAFT [15] results to get a sense of the overall picture.

In addition, Table 1 also shows that our active learning method works favorably against the baseline ("rand"). We found that our active learning method ("occ") may overly sample the same scenes (*e.g.,* "ambush"), so we also test an alternative ("occ-2x") to balance the queried samples. Specifically, we select a double number of samples with top uncertainties and then randomly sample a half from them to query labels. This helps diversify our selected samples when the label ratio is very small. Our active learning methods perform comparably or better than the baseline, especially on the realistic final pass.

Table 2 shows our benchmark testing results on KITTI. Consistent with our findings on Sintel, our semi-supervised methods are significantly better than the compared unsupervised state-of-the-art methods, and close to the supervised IRR-PWC [15], even if we only use a very small fraction (5–20%) of labels. In addition, our active learning method also works consistently better than the baseline for all tested label ratios, especially on the harder KITTI-2015 set.

Ablation Study on Settings of Stage 2. We try different active learning schedules in Stage 2 and show our current setting works the best. We report the Sintel final EPE for different Stage 2 settings with label ratio $r = 0.1$. In Table 3, the first row is our current Stage 2 setting, *i.e.,* semi-supervised training on the partial labeled train set. The second row refers to supervised training only on the labeled part of train set, without the unsupervised samples. The third row considers also including the unlabeled raw data (used in Stage 1) in the Stage 2 semi-supervised training. We can see that our current setting works significantly better than the two alternatives. The second setting works poorly due to overfitting on the very small labeled set, which means that the unlabeled

Table 1. Sintel benchmark results (EPE/px). Metrics evaluated at 'all' (all pixels), 'noc' (non-occlusions), and 'occ' (occlusions). The key metrics (used to sort on the official website) are underlined. Parenthesis means evaluation data used in training. For all metrics, lower is better.

	Label ratio r	Method	Train		Test					
			Clean	Final	Clean			Final		
			all	all	all	noc	occ	all	noc	occ
unsup	$r = 0$	SelFlow [29]	(2.88)	(3.87)	6.56	2.67	38.30	6.57	3.12	34.72
		UFlow [19]	(2.50)	(3.39)	5.21	2.04	31.06	6.50	3.08	34.40
		ARFlow [28]	(2.79)	(3.73)	**4.78**	**1.91**	**28.26**	**5.89**	**2.73**	**31.60**
semi-sup	$r = 0.05$	Ours(rand)	(2.09)	(2.99)	4.04	1.52	24.65	5.49	2.62	28.86
		Ours(occ)	(1.95)	(2.38)	4.11	1.63	**24.39**	**5.28**	**2.49**	**28.03**
		Ours(occ-2x)	(1.94)	(2.55)	**3.96**	**1.45**	24.42	5.35	2.50	28.58
	$r = 0.1$	Ours(rand)	(2.36)	(3.18)	**3.91**	**1.47**	**23.82**	5.21	2.46	27.66
		Ours(occ)	(1.64)	(1.98)	4.28	1.68	25.49	5.31	**2.44**	28.68
		Ours(occ-2x)	(1.75)	(2.30)	4.06	1.63	23.94	**5.09**	2.49	**26.31**
	$r = 0.2$	Ours(rand)	(2.17)	(2.93)	3.89	1.56	**22.86**	5.20	2.50	27.19
		Ours(occ)	(1.35)	(1.63)	4.36	1.86	24.76	5.09	2.45	26.69
		Ours(occ-2x)	(1.57)	(2.05)	**3.79**	**1.44**	23.02	**4.62**	**2.07**	**25.38**
sup	$r = 1$	PWC-Net [47]	(2.02)	(2.08)	4.39	1.72	26.17	5.04	2.45	26.22
		IRR-PWC [15]	(1.92)	(2.51)	3.84	1.47	23.22	4.58	2.15	24.36
		RAFT [49]	(0.77)	(1.27)	**1.61**	**0.62**	**9.65**	**2.86**	**1.41**	**14.68**

part of the train split helps prevent overfitting. The third setting also fails due to the excessive amount of unlabeled data used in Stage 2, which overwhelms the small portion of supervised signal from queried labels.

Model Analysis and Visualization. Table 4(a) shows which Sintel samples are selected by different active learning methods. As shown in the left-most column, the pre-trained model after Stage 1 generally has high EPEs (top 20% shown in the figure) on four scenes, namely "ambush", "cave", "market", and "temple". The random baseline tends to select a bit of every scene, whereas all our three active learning algorithms query scenes with high EPEs for labels. This confirms that our active learning criteria capture samples that are especially challenging to the current model, which explains the success of active learning.

We also analyze the relationships between our criteria and model errors through correlation matrices visualized by heat maps in Figs. 4(b) and 4(c). We can see that the sample errors in Sintel generally have high correlations with all three score values, whereas in KITTI the correlations are much smaller. Also, the "occ ratio" score generally has the highest correlation with sample errors among the three proposed methods. All these observations are consistent with our active

Table 2. KITTI benchmark results (EPE/px and Fl/%). Metrics evaluated at 'all' (all pixels, default for EPE), 'noc' (non-occlusions), 'bg' (background), and 'fg' (foreground). Key metrics (used to sort on the official website) are underlined. '()' means evaluation data used in training. '-' means unavailable. For all metrics, lower is better.

Label ratio r		Method	Train		Test					
			2012	2015	2012		2015			
			EPE	EPE	Fl-noc	EPE	Fl-all	Fl-noc	Fl-bg	Fl-fg
unsup	$r = 0$	SelFlow [29]	(1.69)	(4.84)	4.31	2.2	14.19	9.65	12.68	21.74
		UFlow [19]	(1.68)	(2.71)	4.26	1.9	11.13	8.41	9.78	17.87
		ARFlow [28]	(1.44)	(2.85)	-	1.8	11.80	-	-	-
semi-sup	$r = 0.05$	Ours(rand)	(1.25)	(2.61)	3.90	1.6	9.77	6.99	8.33	17.02
		Ours(occ)	(1.22)	(2.29)	3.90	**1.5**	**9.65**	**6.94**	**8.20**	**16.91**
	$r = 0.1$	Ours(rand)	(1.21)	(2.56)	3.75	1.5	9.51	6.69	8.01	17.01
		Ours(occ)	(1.21)	(1.98)	**3.74**	1.5	**8.96**	**6.28**	**7.74**	**15.04**
	$r = 0.2$	Ours(rand)	(1.16)	(2.10)	3.50	1.5	8.38	**5.68**	7.37	**13.44**
		Ours(occ)	(1.13)	(1.73)	**3.49**	1.5	**8.30**	5.69	**7.25**	13.53
sup	$r = 1$	PWC-Net [47]	(1.45)	(2.16)	4.22	1.7	9.60	6.12	9.66	9.31
		IRR-PWC [15]	-	(1.63)	3.21	1.6	7.65	4.86	7.68	7.52
		RAFT [49]	-	(0.63)	-	-	**5.10**	**3.07**	**4.74**	**6.87**

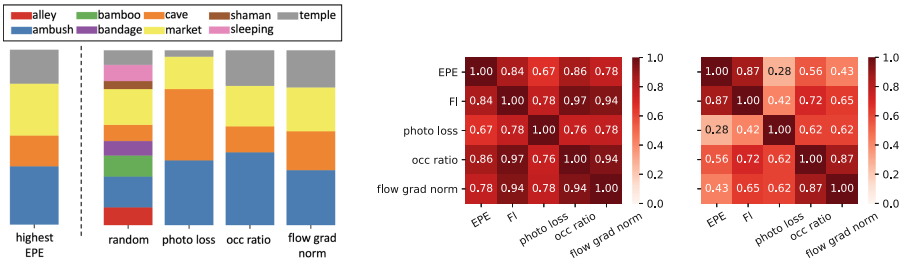
Table 3. Ablation study: different Stage 2 settings. Sintel final validation EPE, label ratio $r = 0.1$. Standard deviations from the last 50 epoches. * denotes current setting.

Data split [Loss]	Method			
	Random	Photo loss	Occ ratio	Flow grad norm
Train [semi-sup]*	**2.71(±0.02)**	**2.54(±0.02)**	**2.52(±0.02)**	**2.54(±0.02)**
Train [sup]	2.82(±0.01)	2.82(±0.01)	2.59(±0.01)	2.77(±0.01)
Raw+train [semi-sup]	3.13(±0.04)	3.09(±0.05)	3.15(±0.06)	3.07(±0.05)

learning validation results. Thus, we posit that the correlation between uncertainty values and sample errors can be a good indicator in designing effective active learning criteria.

Discussion on Factors That May Influence Active Learning

– **Pattern Homogeneity**: Based on our validation results in Fig. 3, active learning seems more effective on Sintel than on KITTI. This may be because KITTI samples are relatively more homogeneous in terms of motion patterns. Unlike the Sintel movie sequences, which contain arbitrary motions of various scales, driving scenes in KITTI exhibit a clear *looming motion* caused by the dominant forward motion of the vehicle that carries the camera. Specif-



(a) Selected samples in Sintel (b) Sintel (c) KITTI

Fig. 4. (a) Sintel samples selected by different methods ($r = 0.2$), grouped by scenes; and correlation matrices with sample errors for Sintel (b) and KITTI (c).

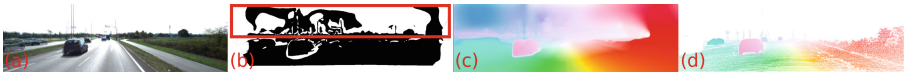


Fig. 5. An example for the information mismatch problem. Data from KITTI-2015, frame 79 (Fl=19.35%), with the third largest "occ ratio" score: (a) superposed input images; (b) estimated occlusion map; (c) flow prediction; (d) flow ground truth.

ically, Sintel has extremely hard scenes like "ambush" as well as extremely easy scenes like "sleeping". This large variation of difficulty makes it possible to select outstandingly helpful samples and labels. In contrast, since KITTI motions are more patterned and homogeneous, any selection tends to make little difference with respect to random sampling.

– **Label Region Mismatch**: KITTI only has sparse labels, *i.e.,* only a part of the image pixels have labels. This is crucial because our active learning criteria are computed over the whole frame, so there is a mismatch between the support of our criteria and the KITTI labels. Specifically, the sparse labels may not cover the problematic regions found by our criteria. One example is shown in Fig. 5. The sky region has bad predictions due to lack of texture, and the "occ ratio" method captures the inconsistent flow there by highlighting the sky region. However, the ground-truth labels do not cover the sky region, so having this sample labeled does not help much in training.

5 Conclusion

In this paper, we first analyzed the trade-off between model performance and label ratio using a simple yet effective semi-supervised optical flow network and found that the unsupervised performance can be significantly improved even with a small fraction of labels. We then explored active learning as a way to further improve the performance and reduce annotation costs. Our active learning method works consistently better than baseline on Sintel and KITTI datasets.

For potential future work, it may be interesting to explore how to deal with sparse labels in the active learning framework or how to query labels by region rather than full frame.

Acknowledgments. This material is based upon work supported by the National Science Foundation under Grant No. 1909821 and by the Intelligence Advanced Research Projects Agency under contract number 2021-21040700001.

References

1. Aslani, S., Mahdavi-Nasab, H.: Optical flow based moving object detection and tracking for traffic surveillance. Int. J. Elect. Comput. Energ. Electron. Commun. Eng. **7**(9), 1252–1256 (2013)
2. Beluch, W.H., Genewein, T., Nürnberger, A., Köhler, J.M.: The power of ensembles for active learning in image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9368–9377 (2018)
3. Butler, D.J., Wulff, J., Stanley, G.B., Black, M.J.: A naturalistic open source movie for optical flow evaluation. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7577, pp. 611–625. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33783-3_44
4. Choi, J., Elezi, I., Lee, H.J., Farabet, C., Alvarez, J.M.: Active learning for deep object detection via probabilistic modeling. In: Proceedings of the IEEE International Conference on Computer Vision (2021)
5. Dosovitskiy, A., et al.: FlowNet: learning optical flow with convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2758–2766 (2015)
6. Ebrahimi, S., Elhoseiny, M., Darrell, T., Rohrbach, M.: Uncertainty-guided continual learning with Bayesian neural networks. In: International Conference on Learning Representations (2020)
7. Fan, L., Huang, W., Gan, C., Ermon, S., Gong, B., Huang, J.: End-to-end learning of motion representation for video understanding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6016–6025 (2018)
8. Gal, Y., Islam, R., Ghahramani, Z.: Deep Bayesian active learning with image data. In: Proceedings of the International Conference on Machine Learning, pp. 1183–1192. PMLR (2017)
9. Gao, C., Saraf, A., Huang, J.-B., Kopf, J.: Flow-edge guided video completion. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12357, pp. 713–729. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58610-2_42
10. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: the kitti dataset. Int. J. Robot. Res. **32**(11), 1231–1237 (2013)
11. Grandvalet, Y., Bengio, Y.: Semi-supervised learning by entropy minimization. In: Advances in Neural Information Processing Systems, vol. 17. MIT Press (2005)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
13. Houlsby, N., Huszár, F., Ghahramani, Z., Lengyel, M.: Bayesian active learning for classification and preference learning. arXiv preprint arXiv:1112.5745 (2011)

14. Hui, T.W., Tang, X., Change Loy, C.: LiteflowNet: a lightweight convolutional neural network for optical flow estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8981–8989 (2018)
15. Hur, J., Roth, S.: Iterative residual refinement for joint optical flow and occlusion estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5754–5763 (2019)
16. Ilg, E., Saikia, T., Keuper, M., Brox, T.: Occlusions, motion and depth boundaries with a generic network for disparity, optical flow or scene flow estimation. In: Proceedings of the European Conference on Computer Vision, pp. 614–630 (2018)
17. 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, pp. 690–706 (2018)
18. Yu, J.J., Harley, A.W., Derpanis, K.G.: Back to basics: unsupervised learning of optical flow via brightness constancy and motion smoothness. In: Hua, G., Jégou, H. (eds.) Computer Vision – ECCV 2016 Workshops. LNCS, vol. 9915, pp. 3–10. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-49409-8_1
19. Jonschkowski, R., Stone, A., Barron, J.T., Gordon, A., Konolige, K., Angelova, A.: What matters in unsupervised optical flow. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12347, pp. 557–572. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58536-5_33
20. Kapoor, A., Grauman, K., Urtasun, R., Darrell, T.: Active learning with gaussian processes for object categorization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1–8. IEEE (2007)
21. Kim, H.H., Yu, S., Tomasi, C.: Joint detection of motion boundaries and occlusions. In: British Machine Vision Conference (2021)
22. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: International Conference on Learning Representations (2014)
23. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
24. Lai, W.S., Huang, J.B., Yang, M.H.: Semi-supervised learning for optical flow with generative adversarial networks. In: Advances in Neural Information Processing Systems, pp. 353–363 (2017)
25. Laine, S., Aila, T.: Temporal ensembling for semi-supervised learning. In: International Conference on Learning Representations (2017)
26. Lee, D.H., et al.: Pseudo-label: the simple and efficient semi-supervised learning method for deep neural networks. In: Workshop on challenges in representation learning, ICML, vol. 3, p. 896 (2013)
27. Li, X., Guo, Y.: Adaptive active learning for image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 859–866 (2013)
28. Liu, L., et al.: Learning by analogy: Reliable supervision from transformations for unsupervised optical flow estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6489–6498 (2020)
29. Liu, P., Lyu, M., King, I., Xu, J.: Selflow: self-supervised learning of optical flow. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4571–4580 (2019)
30. Liu, Y., Chen, K., Liu, C., Qin, Z., Luo, Z., Wang, J.: Structured knowledge distillation for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2604–2613 (2019)

31. Mackowiak, R., Lenz, P., Ghori, O., Diego, F., Lange, O., Rother, C.: Cereals-cost-effective region-based active learning for semantic segmentation. In: British Machine Vision Conference (2018)
32. Mayer, N., Ilg, E., Häusser, 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 International Conference on Computer Vision (2016). arXiv:1512.02134
33. 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 (2018)
34. Menze, M., Geiger, A.: Object scene flow for autonomous vehicles. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015)
35. Miyato, T., Maeda, S.I., Koyama, M., Ishii, S.: Virtual adversarial training: a regularization method for supervised and semi-supervised learning. IEEE Trans. Pattern Anal. Mach. Intell. **41**(08), 1979–1993 (2019)
36. Paszke, A., et al.: Pytorch: an imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems, pp. 8024–8035. Curran Associates, Inc. (2019)
37. Paul, S., Bappy, J.H., Roy-Chowdhury, A.K.: Non-uniform subset selection for active learning in structured data. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6846–6855 (2017)
38. Ranjan, A., Black, M.J.: Optical flow estimation using a spatial pyramid network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4161–4170 (2017)
39. Ranjan, A., et al.: Competitive collaboration: Joint unsupervised learning of depth, camera motion, optical flow and motion segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 12240–12249 (2019)
40. 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 (2017)
41. Ren, Z., Gallo, O., Sun, D., Yang, M.H., Sudderth, E.B., Kautz, J.: A fusion approach for multi-frame optical flow estimation. In: Winter Conference on Applications of Computer Vision, pp. 2077–2086. IEEE (2019)
42. Roy, S., Unmesh, A., Namboodiri, V.P.: Deep active learning for object detection. In: British Machine Vision Conference, vol. 362, p. 91 (2018)
43. Sajjadi, M., Javanmardi, M., Tasdizen, T.: Regularization with stochastic transformations and perturbations for deep semi-supervised learning. Adv. Neural. Inf. Process. Syst. **29**, 1–10 (2016)
44. Siddiqui, Y., Valentin, J., Nießner, M.: Viewal: Active learning with viewpoint entropy for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9433–9443 (2020)
45. Song, X., Zhao, Y., Yang, J., Lan, C., Zeng, W.: FPCR-net: feature pyramidal correlation and residual reconstruction for semi-supervised optical flow estimation. arXiv preprint arXiv:2001.06171 (2020)
46. 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 Conference on Computer Vision and Pattern Recognition, pp. 3887–3896 (2021)
47. 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)

48. Tarvainen, A., Valpola, H.: Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results. Adv. Neural. Inf. Process. Syst. **30**, 1–10 (2017)
49. Teed, Z., Deng, J.: RAFT: recurrent all-pairs field transforms for optical flow. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12347, pp. 402–419. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58536-5_24
50. K. Z. Tehrani, A., Mirzaei, M., Rivaz, H.: Semi-supervised training of optical flow convolutional neural networks in ultrasound elastography. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12263, pp. 504–513. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59716-0_48
51. Wang, R., Wang, X.Z., Kwong, S., Xu, C.: Incorporating diversity and informativeness in multiple-instance active learning. IEEE Trans. Fuzzy Syst. **25**(6), 1460–1475 (2017)
52. 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)
53. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
54. Wei, K., Iyer, R., Bilmes, J.: Submodularity in data subset selection and active learning. In: Proceedings of the International Conference on Machine Learning, pp. 1954–1963. PMLR (2015)
55. Xie, Q., Luong, M.T., Hovy, E., Le, Q.V.: Self-training with noisy student improves imagenet classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 10687–10698 (2020)
56. Yan, W., Sharma, A., Tan, R.T.: Optical flow in dense foggy scenes using semi-supervised learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 13259–13268 (2020)
57. Yang, Y., Soatto, S.: Conditional prior networks for optical flow. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11219, pp. 282–298. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01267-0_17
58. Yin, Z., Shi, J.: GeoNet: unsupervised learning of dense depth, optical flow and camera pose. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1983–1992 (2018)
59. Yu, H., Chen, X., Shi, H., Chen, T., Huang, T.S., Sun, S.: Motion pyramid networks for accurate and efficient cardiac motion estimation. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12266, pp. 436–446. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59725-2_42
60. Zhang, F., Woodford, O.J., Prisacariu, V.A., Torr, P.H.: Separable flow: learning motion cost volumes for optical flow estimation. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 10807–10817 (2021)