Optical Flow Models and Training Techniques in Data-Constrained Environment

Senem Isik Shawn Zixuan Kang Zander Lack

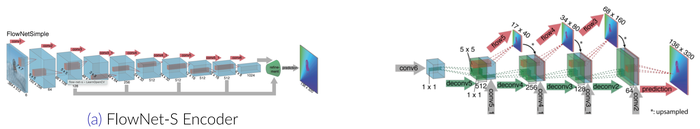
Stanford University

Introduction

Optical flow estimation aims at calculating a pixel-wise displacement field between two images. The flow field provides crucial correspondence information that has significant applications to tasks such as image interpolation, super-resolution reconstruction, object segmentation and tracking, action recognition, and autonomous navigation.

Background: Optical Flow Estimation Models

Architectures



(a) FlowNet-S Encoder

(b) FlowNet-S Decoder

Figure 1. FlowNet-S architecture

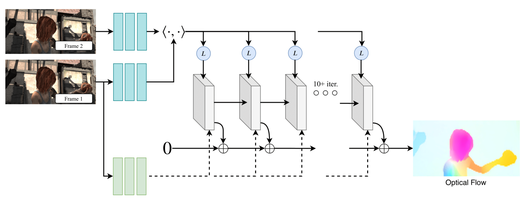


Figure 2. RAFT architecture

Loss Function

$$EPE = ||V_{gt} - V_{calc}||_2 =$$

$$\sqrt{(\Delta x_{gt} - \Delta x_{calc})^2 + (\Delta y_{gt} - \Delta y_{calc})^2}$$

Problem Statement

One current limitation of optical flow research is limited data. Getting optical flow ground truth for realistic video material is difficult. There are only a few datasets and all of the larger datasets are generated with computer simulation techniques. We will analyse the effectiveness of existing optical flow architectures specifically FlowNet and RAFT in a data-constrained environment, and use those insights to modify an existing architecture to hopefully improve its performance. Specifically, we will experiment with data augmentation and different dataset sizes to determine the effectiveness of each model.

Dataset

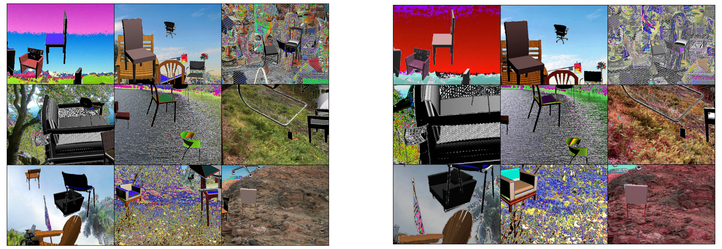
We used synthetic dataset Flying Chairs and MPI Sintel to train and test the models and realistic dataset KITTI for additional testing.

Methods

New Architecture: Deconvolutional Upsampling for FlowNet

The prediction process in Flownet traditionally involves taking the decoder output, passing it through an additional convolutional layer and then using bilinear upsampling. Our change combines these two last steps, taking the decoder output and applying a transposed convolution to simultaneously extract features and upsample the output. Hopefully, this will provide a more granular predicted flow that can achieve better performance as it uses trainable weights to upsample the flow.

Data Augmentation



(a) Original Samples

(b) Samples with Data Augmentation

Figure 3. Data Augmentation on Flying Chairs Dataset

Experiments

To explore the performance of the models with different training methods, we pre-trained FlowNet and RAFT with 0, 500, and 2000 image pairs from Flying Chairs and fine-tuned on 520 image pairs from Sintel, with and without data augmentation respectively.

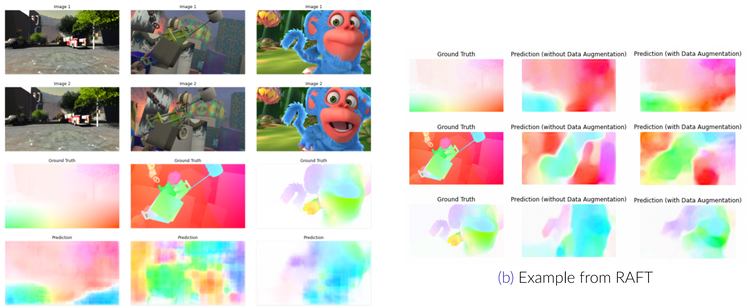
Results: Testing FlowNet with the New Architecture

The architecture change yields no significant change in performance from the model. Two models achieve nearly identical scores, though the new architecture does slightly better with less data and the traditional FlowNet does slightly better with more pre-training data.

Table 1. Evaluating FlowNet with a new Upsampling Architecture on Sintel

Size	Model	EPE	1px	3px	5px	F1
0	New	**9.28**	**27.16**	**55.44**	**68.19**	**44.56**
	Old	10.18	23.27	51.73	63.36	48.27
500	New	7.88	**46.52**	**67.96**	75.48	32.03
	Old	7.88	46.42	68.29	**75.81**	**31.70**
2000	New	**6.80**	54.25	73.38	79.65	26.57
	Old	6.79	**54.39**	**73.79**	**79.88**	**26.18**

Results: Comparing RAFT and FlowNet in Data-Constraint Environment



(a) Example from FlowNet

(b) Example from RAFT

Figure 4. Results from FlowNet and RAFT pre-trained and fine-tuned in limited datasets

Evaluating Flownet on Sintel

Size	DA	EPE	1px	3px	5px	F1
0	no	10.18	23.27	51.73	63.36	48.27
	yes	10.43	30.72	56.04	67.26	43.97
500	no	7.88	46.42	68.29	75.81	31.70
	yes	7.74	37.49	65.50	75.03	34.48
2000	no	6.79	**54.39**	**73.79**	**79.88**	**26.18**
	yes	**6.47**	47.69	71.24	79.03	28.72

Evaluating RAFT on Sintel

Size	DA	EPE	1px	3px	5px	F1
0	no	45.02	3.45	20.16	33.59	79.83
	yes	18.37	3.24	23.91	40.29	78.09
500	no	28.67	8.48	29.33	37.57	70.66
	yes	8.70	15.73	59.57	75.26	40.38
2000	no	38.78	14.39	32.51	42.54	67.49
	yes	**6.51**	**44.73**	**80.59**	**87.05**	**19.15**

Table 2. Evaluating models in different training methods. DA stands for data augmentation.

Conclusion and Future Work

From our experiments, we observed that

- Data size is crucial to training optical flow models in data-constrained environments
- Data augmentation helps models generalize better, however, it improves RAFT significantly more than it improves FlowNet

For future work:

- Conduct similar experiments on different optical flow models (PWC-net, IRR, etc.) to explore how different models perform in data-constrained environments and how conditions including data augmentation and training data size effect their performance.
- Single out and implement different aspects of the latter models on FlowNet to understand which feature is the most crucial for improving optical flow model performances under data-constrained environments