Review for RAFT: Recurrent All-Pairs Field Transforms for Optical Flow

Irem Arpag

March 14, 2021

The paper focuses on one of the most significant but outstanding problem of computer vision "optical flow" that is a challenging task because of difficulties such as occlusions, fast moving objects, textureless areas and motion blur. Since traditional optical flow models show limited improvement because of hand-design optimization approaches, recent deep learning methods, one of them is "Recurrent All-Pairs Field Transform" called shortly "RAFT" introduced by the authors of the paper, have the advantage of training a network to directly predict flow resulted in fast inference time. With its innovative three sectioned architecture composed of a feature encoder that extracts a feature vector for each pixel; a correlation layer that designs a 4D correlation volume for all pairs of pixels; an update operator which regains values from the correlation volumes, RAFT is claiming extra challenges on effective performance, generalization and easy training comparing to current deep learning methods. Although the model takes inspiration from the a/m traditional optimization based approaches, its significance comes from not computed manually but learned through training by the encoder and operator components separately.

The quantitative results of the experiments on KITTI and SINTEL indicate the strength of the model on accuracy with its performance 16 on KITTI, and 30 on SINTEL error reduction percentage from the best published results; on generalization with its achievement of 40 in percentage error reduction from the best trained network on KITTI synthetic dataset; and on efficiency by training with 10x fewer iterations than prior networks.

"RAFT" is evaluated mainly by using KITTI and SINTEL datasets and pretrained on FlyingChairs (C) and FlyingThings (T) for finetuning, also tested by using HD video from DAVIS dataset to confirm its high-resolution capacity. The results of the experiments made by using KITTI ranks first on the leaderboard performing a deduction of 3.32 (from 8.36 to 5.04) EPE (End Point Error) that demonstrates generalizations capacity of the model. The performance of the architecture on SINTEL after pretraining on C+T is 29 in percentage lower than FlowNet2 with an average EPE of 1.43 indicating good cross dataset generalization. Comparing the model with a group of optical flow deep-learning methods by taking parameters from their papers shows that "RAFT" is more efficient in terms of inference time, parameter count, and training iterations. In ablation studies, each component of the approach (feature encoder, correlation layer, update operator) Is tested in isolation that demonstrate the Importance of each part separately.

RAFT, as a deep learning method solving optical flow iteratively, got the best paper award of ECCV 2020.[1] From the source provided by the authors, some researchers tested this state-of-art model with their own video clips and the visualization results seem satisfactory. As a final note, a current article dated Feb 2021 [2] proposing a joint unsampling approach for estimating optical flow performs better results than "RAFT" both on KITTI and SINTEL benchmarks.

REFERENCES

[1] Shuchen DU, Understanding Optical Flow and RAFT, 2020, Towards Data Science

[2] Eldesokey A, Felsberg M, Normalized Convolution Unsampling for Refined Optical Flow Estimation Feb 2021. Arxiv