Learning Optical Flow with Deformable Convolution

Group 1
Cheng-Wei Chen, Cheng Chien, Pin-Chen Huang

Introduction

Optical flow is a fundamental method used for quantitative motion estimation on the image plane. In the deep learning era, most works treat it as a task of “matching of features”, learning to pull matched pixels as close as possible in feature space and vice versa. However, spatial affinity (smoothness constraint), another important component for motion understanding, has been largely overlooked.

First, we implemented the feature extraction in “RAFT: Recurrent All-Pairs Field Transforms for Optical Flow”[1] . In this way, we can generate 4D-CV(4D Correlation Volume) by calculating two feature maps extracted from two consecutive frames. Furthermore, “Learning optical flow with kernel patch attention”[3] , introduces a novel approach, called kernel patch attention (KPA), to better resolve the ambiguity in dense matching by explicitly considering the local context relations. Similarly, “Deformable Convolutional Neural Networks” [2]also concentrates on local context. Both deformable convolution and deformable RoI(region-of-interest) pooling are based on the idea of augmenting the spatial sampling locations in the modules with additional offsets and learning the offsets from the target tasks, without additional supervision.

By the combination of the above methods, we believe the result can be as great as using KPA(Kernel Patch Attention) module[3] . On two different datasets (Sintel and Flying chairs), corresponding to consecutive sequences and a pair of two continuous frames, our proposed deformable flow achieves the best performance with an EPE of 3.13 on the clean pass.

Methodology

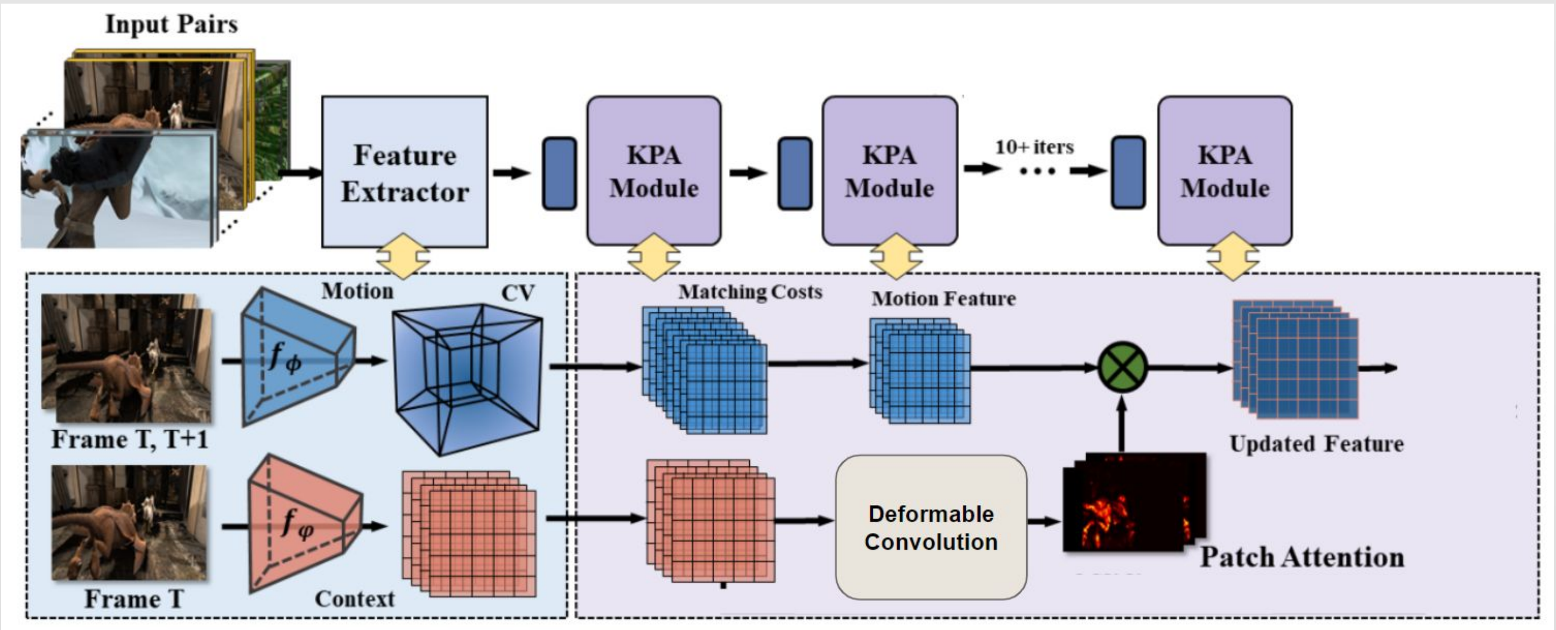


Figure 1. Architecture of our model

We first refer to “RAFT: Recurrent All-Pairs Field Transforms for Optical Flow”[1] , using two encoders to extract a pair of feature maps(f_1, f_2) and context feature, then we build a 4D correlation volume based on f_1 and f_2.

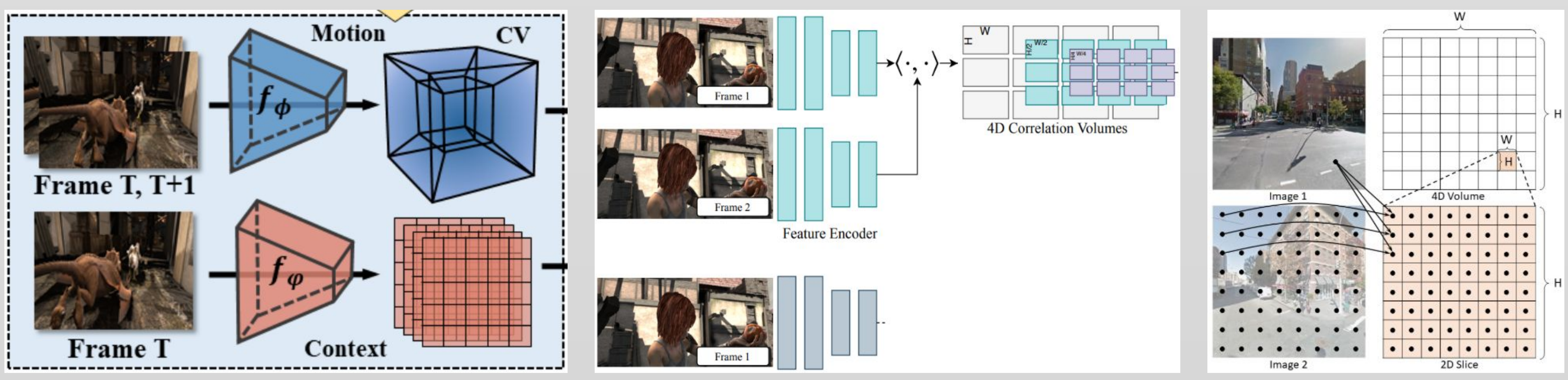


Figure 2. Details of feature extractor and 4D-CV

After deriving the 4D CV, we capture the motion feature f_m by applying a motion encoder on the segmented matching costs. We build the encoder based on six residual blocks with three strides, and the channel dimension of the output feature map is set to 256.

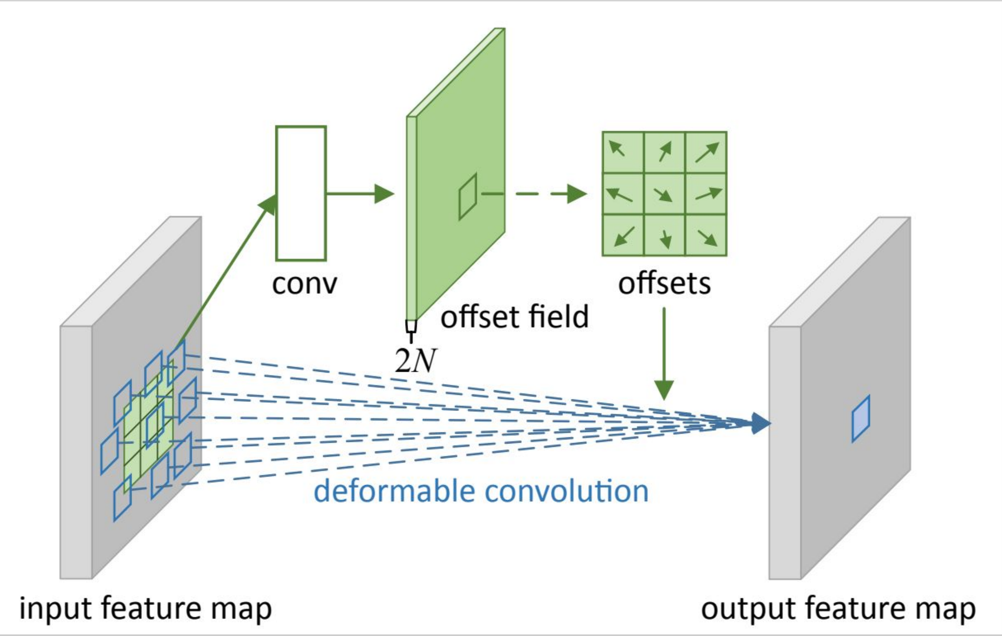
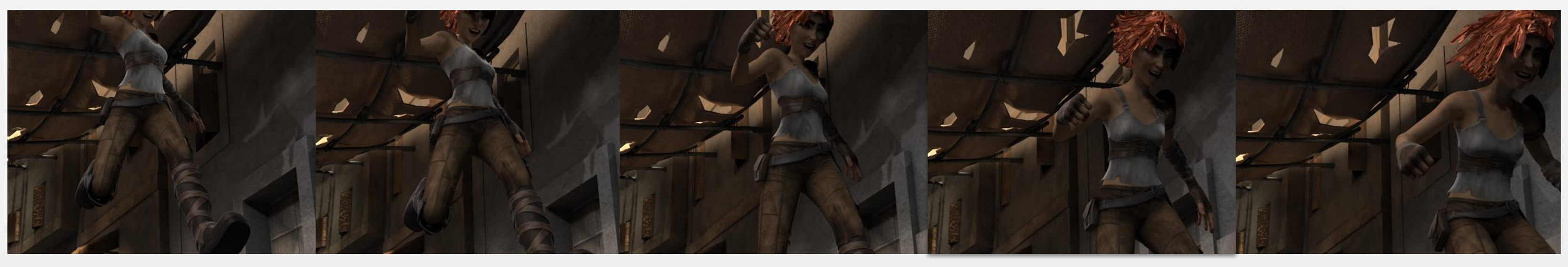


Figure 3. Architecture of deformable convolution

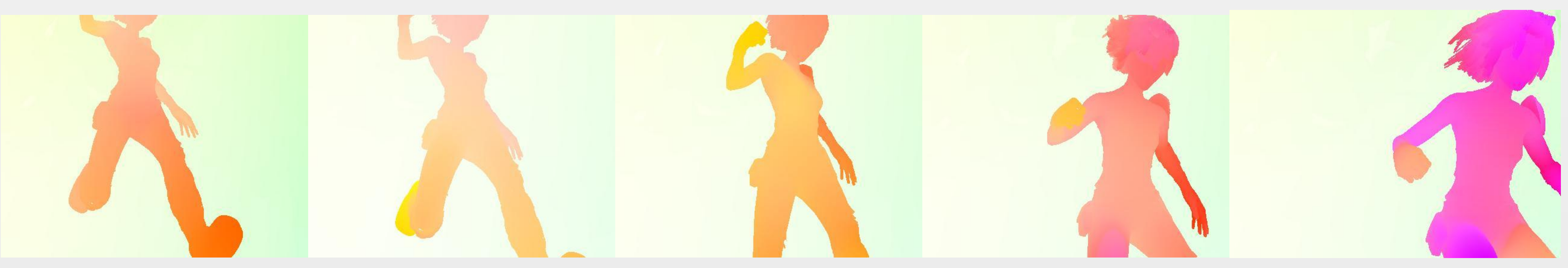
Moreover, if we input the context feature to the deformable convolutional layers[2], it will generate the output feature map by using another learnable offsets map. Then, combine the motion feature and context feature map to update our new motion feature. Finally, concatenate the motion feature with the original context feature and decode it to an optical flow.

Results

We implement deformable convolution[2] to generate the optical flow and experiments based on PyTorch toolbox. Fig. 4 provides that applied deformable convolution[2] method able to focus on local and boundary information. First, we trained for 60 epoch and using FlyingChair as dataset. Second, we used the weight trained by Flying-Chair as the pre-train weight, and keep training 100 epochs by using Sintel. In Fig 5. we can observe that the end-point-error(EPE) can reduce to about 3.13.



(a) Reference Frame



(b) Ground Truth Flow

(c) Predict Truth Flow

Figure 4. Final results

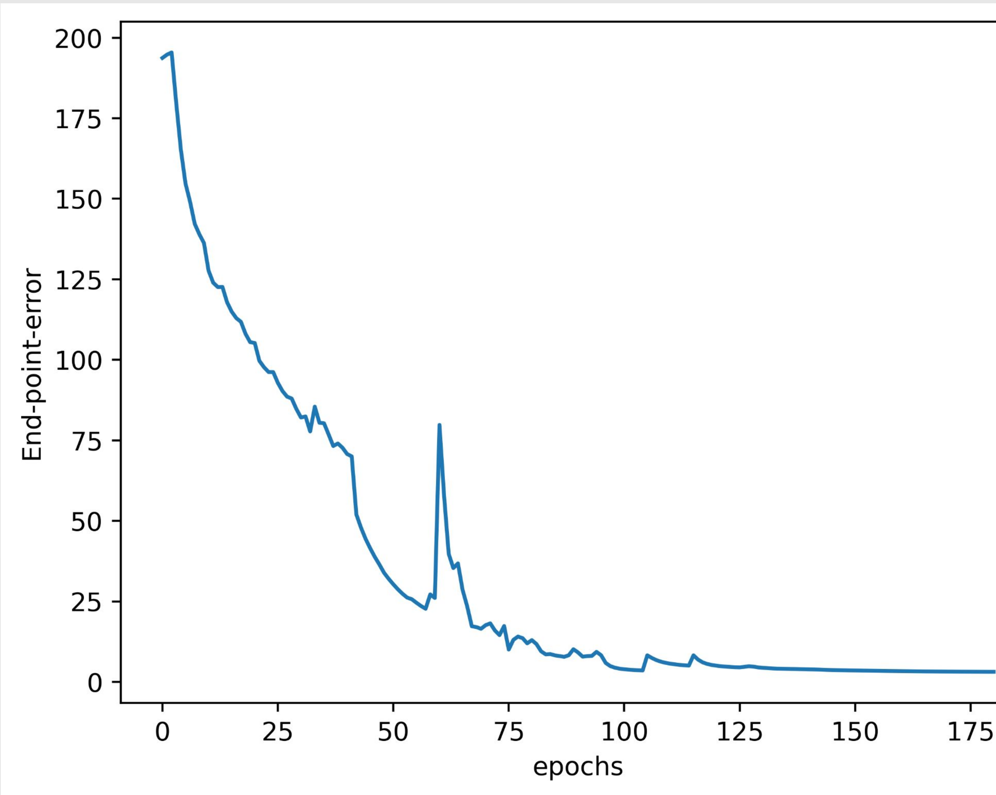


Figure 5. End-Point-Error diagram

Conclusion

Our experiment result shows that it is not as good as we expect. The best EPE on Sintel Dataset(clean pass) is 3.13, we believe that we can review and improve the model. Improvement and future works are as follow:

1. Iterate deformable module(new KPA flow) more times.
2. Review the process of mapping context domain to attention domain.
3. Combine two benefits of deformable convolution and kernel patch attention.

References

[1] Zachary Teed, Jun Deng. “Raft: Recurrent all-pairs field transforms for optical flow.”, ECCV 2020.

[2] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, Yichen Wei, “Deformable Convolutional Networks”, CVPR 2017.

[3] Ao Luo, Fan Yang, Xin Li, Shuaicheng Liu, “Learning Optical Flow With Kernel Patch Attention”, CVPR 2022.

[4] Ao Luo, Fan Yang, Kunming Luo, Xin Li, Haoqiang Fan, and Shuaicheng Liu. Learning optical flow with adaptive graph reasoning. In AAAI, 2022.

[5] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He, “Non-local neural networks”, CVPR 2018.