Your Objective Is Wrong: Rethink Unsupervised learning of Optical Flow

Anonymous CVPR submission

Paper ID ****

Abstract

The ABSTRACT is to be in fully-justified italicized text, at the top of the left-hand column, below the author and affiliation information. Use the word "Abstract" as the title, in 12-point Times, boldface type, centered relative to the column, initially capitalized. The abstract is to be in 10-point, single-spaced type. Leave two blank lines after the Abstract, then begin the main text. Look at previous CVPR abstracts to get a feel for style and length.

1. Introduction

1.1. Related Work

1.2. Novel Contribution

We extend FlowNet [2] in this work, in summary our contributions are three folds. First, we proposed to combine traditional layered approach for optical flow estimation with deep learning. The proposed approach does not require pre-segmentation of images, instead, the separation of layers is automatically done during training the network. Second, a soft-masks module is proposed. This soft-masks module implements a channel-wise maxout operation among masks. As a result, the estimated optical flow will be separated to layers. each of which will contain optical flow that is estimated using a quadratic function. Third, we extend the FlowNet by adding the proposed soft-mask module in output layers, the resulting network is trained to compare with both supervised and unsupervised optical flow estimation approaches using neural networks. The empirical results show that the proposed network structure achieves comparable or lower error in each experimental group.

2. The Proposed Approach

Given a pair of images $I_a, I_b \in \mathbb{R}^{H\times W\times C}$ as input, where H, W and C are height, width and channels of the input images. The proposed approach is going to estimate an optical flow field $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{H\times W}$, which defines a transformation from I_a to I_b. In this section, first we are going to illustrate mistakes made by existing unsupervised optical flow neural networks. A solution that provides a correct objective function then is proposed. We highlight the details of the proposed optical flow estimation neural network in the end.

2.1. Correct Objective Function

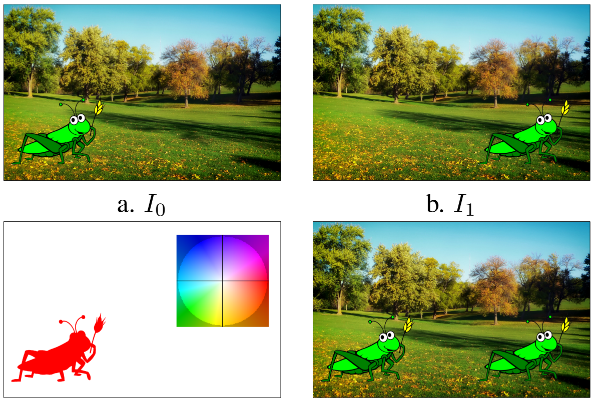
For existing works[1][4][5], unsupervised learning of optical flow is achieved by using a network framework that combines FlowNet [2] and Spatial Transformation Network (STN) [3]. Given a fact that image transformation is a reverse mapping in STN, so, the optical flow is actually transforming I_1 in this case. Existing works follow a same pattern and define their objective function as following:

$$E = \sum_{i}^{H}\sum_{j}^{W}(I_1(i+u_{ij}, j+v_{ij}) - I_0(i,j))^2 + \varphi(\mathbf{u}, \mathbf{v}) \quad (1)$$

where $\varphi((u, v))$ is a regularization term that constrains smoothness of optical flow.

It is helpful to think about a fact that how optical flow exactly transforms image I_1 in reverse mapping in Eqn.1. Suppose a new image denoted as I^* will be produced by transforming I_1 using the optical flow. The expression $I_1(i+u_{ij}, j+v_{i,j})$ in Eqn.1 actually means the pixel $I_1(i+u_{ij}, j+v_{ij})$ will will get *copied* and *moved* by flow vector (u_{ij}, v_{ij}) to pixel $I^*(i,j)$ in image I^*. Then one question raises up that what happens to pixel $I^*(i+u_{ij}, j+v_{ij})$ in image I^*, is it gauranteed that there is always a pixel in image I_0 will get copied to $I^*(i+u_{ij}, j+v_{ij})$? The answer actually depends on optical flows at other pixels. For better understanding, an illustration showing the transformation caused by optical flow is given in Fig.1.

In Fig.1, a grasshopper is going to hop from left in I_0 to right in I_1 with static background in both images. Fig.1c shows the ground truth optical flow from I_0 to I_1, in which only area of grasshopper gets non-zero flow. Now, let us suppose I^* starts as an empty image and we first focus on the area of grasshopper where optical flow is not zero. Given a pixel location (i, j) in the grasshopper area and a



a. I_0 b. I_1 c. Flow d. I^*

Figure 1. Illustration of image transformation as a result of reverse mapping using optical flow.

flow vector at this pixel (u_{ij}, v_{ij}), because of transformation, we will have $I^*(i,j) = I_1(i + u_{ij}, j + v_{ij})$, which moves grasshopper from right side in I_1 to left side in I^*. Then let us consider background areas with no optical flow, following the same rule, we will have $I^*(i,j) = I_1(i+0, j+0)$. That is to say, the background area in I_1 including the grasshopper will be exactly copied to I^*. As a result, we are able to see two grasshoppers in I^*.

It could be known from this illustration that, it is not necessarily true that I^* will be identical to I_0 even being transformed using true optical flow. Using Eqn.1 as an objective thus will not gaurantee that network is able to learn a correct optical flow. Because using I_0 as target image, the training process will force optical flow to be generated to produce I^* which is exactly the same to I_0. As a result, the estimation of optical flow could easily be wrong. So, in this paper, we argue that, under the unsupervised optical flow learning framework, using I_0 as reference image is wrong. Instead the correct reference image is I^*, the learning of which requires accessing to the ground truth optical flow.

2.2. Target Image Estimation Network

2.3. Flow Estimation Network

3. Empirical Evaluation

3.1. Datasets

3.2. Training Details

3.3. Results

4. Discussion and Summary

References

[1] A. Ahmadi and I. Patras. Unsupervised convolutional neural networks for motion estimation. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 1629–1633. IEEE, 2016. 1

[2] A. Dosovitskiy, P. Fischery, E. Ilg, P. Husser, C. Hazirbas, V. Golkov, P. v. d. Smagt, D. Cremers, and T. Brox. Flownet: Learning optical flow with convolutional networks. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 2758–2766, Dec 2015. 1

[3] M. Jaderberg, K. Simonyan, A. Zisserman, and k. kavukcuoglu. Spatial transformer networks. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 2017–2025. Curran Associates, Inc., 2015. 1

[4] Z. Ren, J. Yan, B. Ni, B. Liu, X. Yang, and H. Zha. Unsupervised deep learning for optical flow estimation. In *Artificial Intelligence (AAAI-17), Proceedings of the Thirty-First AAAI Conference on*, pages 1495–1501, 2017. 1

[5] J. J. Yu, A. W. Harley, and K. G. Derpanis. Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness. *CoRR*, abs/1608.05842, 2016. 1