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. Methodology

The proposed approach and corresponding analysis will be introduced in this section.

2.1. Annotation

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 $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{H\times W}$, where $\mathbf{u}$ and $\mathbf{v}$ are the horizontal and vertical components of the optical flow field to be estimated that transform image from I_a to I_b. The original formulation of optical flow estimation is proposed by Horn and Schunck in [3]. In classical formulation, an objective function is composed with a combination of a data term which makes a local constancy assumption of some image property and a spatial term that models how the flow is expected to vary across images. We write the classical optical flow objective function as:

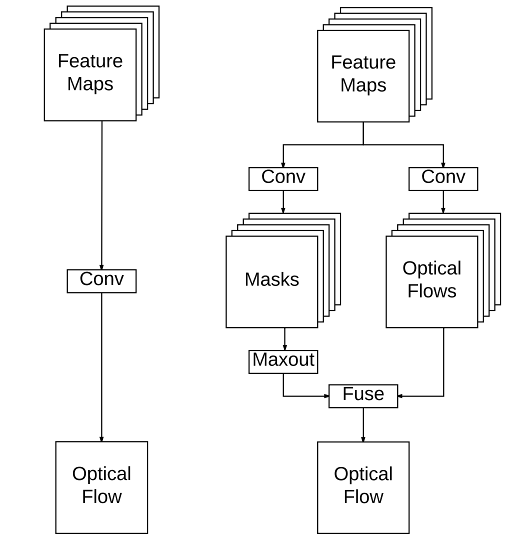
$$E(\mathbf{u},\mathbf{v}) = \sum_{i}^{H}\sum_{j}^{W}(I_1(i+u_{ij}, j+v_{ij}) - I_0(i,j))^2 + \varphi(\mathbf{u},\mathbf{v}) \tag{1}$$

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

Nowadays, the above objective is still being used by many optical flow estimation using deep neural network based on unsupervised training framework [1][4][5]. We also use above objective when training our network and comparing with results of unsupervised methods. Experiments results are presented in Section 3.

2.2. Soft-masks module

FlowNet [2] is the first work that uses deep convolutional neural network for flow estimation. The network architecture used by FlowNet is very similar to classical structure of auto-encoder, where optical flows are generated using deconvolution at each scale level of the image pyramid. To refine estimation of the flows, shortcuts are built to connect layers of corresponding level in encoder and decoder. Given $x_{ij} \in \mathbb{R}^{s\times s\times c}$, where s is kernel size and c is number of channels, representing a feature vector inputted to output layer. FlowNet employs a linear activation to compute optical flow:



a. Normal output b. Soft-masks module

Figure 1. An illustration of the structure of the proposed soft-masks module compared with normal linear optical flow output.

$$f_{ij} = x_{ij}^T W + b \tag{2}$$

Given actual optical flow field a nonlinear and piece-wise smooth representation of motions contained in images, using linear function to fit the flow field is less accurate. We introduce a soft-masks module in this paper. The proposed module can be used to replace the linear output of optical flow estimation. We will show that by using this module, we are able to separate optical flow field to multiple layers. Flow estimation in each layer is smooth and can be estimated using a quadratic function, which results in a more accurate and consistent optical flow estimation.

An illustration of soft-masks module is shown in Figure 1. The essential part of the soft-masks module is its dual-branch design.

2.3. Target Image Estimation Network

2.4. 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] B. K. Horn and B. G. Schunck. Determining optical flow. *Artificial intelligence*, 17(1-3):185–203, 1981. 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