Devon: Deformable Volume Network for Learning Optical Flow

Yao Lu, Jack Valmadre, Heng Wang, Juho Kannala, Mehrtash Harandi, Philip Torr

ANU & Oxford & Facebook & Aalto

November 25, 2023
Optical Flow

Pixel-wise motion between two images
Learning Optical Flow

Why?
- Learn good features for matching
- Scene reasoning (e.g. occlusion, segmentation, semantics)
Large Motion

Standard CNNs only have small receptive field

Standard solution:
*multi-resolution + warping*
Multi-resolution + Warping

Coarse-to-fine optical flow estimation

Gaussian pyramid of image H

Gaussian pyramid of image I

run iterative L-K
warp & upsample
run iterative L-K
Multi-resolution + Warping

Problems:

- small things move fast
- warping
Small Things Move Fast

- At low resolution, small things disappear.
- At high resolution, the motion is too large to be covered.
The Problem of Warping

(a) First image

(b) Second image

(c) Ground truth flow

(d) Warped second image
The Problem of Warping

\[ \tilde{J}(x) = J(x + F(x)) \]

\[ F(p_1) = 0, \quad F(p_2) = p_1 - p_2, \]  \hspace{1cm} (1)

we have

\[ \tilde{J}(p_1) = J(p_1 + F(p_1)) \]
\[ = J(p_1 + 0) = J(p_1), \]  \hspace{1cm} (2)

\[ \tilde{J}(p_2) = J(p_2 + F(p_2)) \]
\[ = J(p_2 + p_1 - p_2) = J(p_1). \]  \hspace{1cm} (5)

\[ \tilde{J}(p_1) = \tilde{J}(p_2) = J(p_1) \]
Solution

Key idea: instead of deforming the images (downsample + warping), we deform the correspondences.

(e) Standard cost volume

(f) Deformable cost volume
Devon

- No multi-resolution
- No warping
- Only 1M parameters
Results

(g) First image

(h) Second image

(i) PWC-Net

(j) LiteFlowNet

(k) Devon

(l) Ground truth