Devon: Deformable Volume Network for Learning Optical Flow

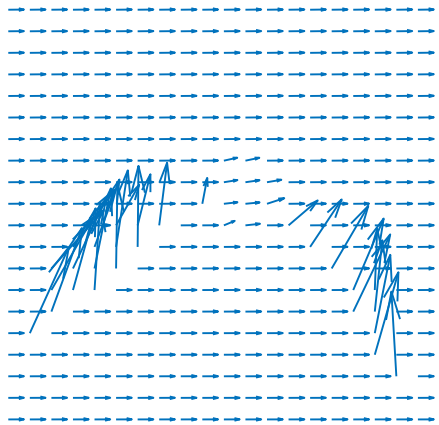
Yao Lu, Jack Valmadre, Heng Wang,
Juho Kannala, Mehrtash Harandi, Philip Torr

ANU & Oxford & Facebook & Aalto

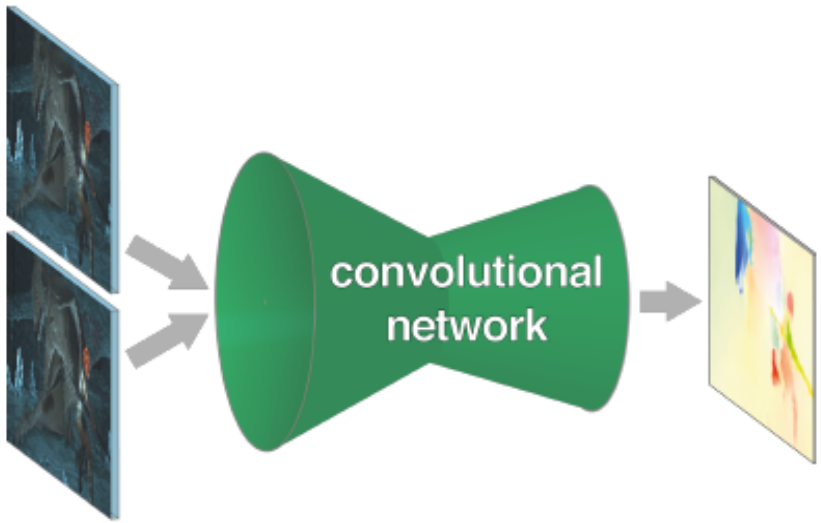
November 25, 2023

Optical Flow

Pixel-wise motion between two images



Learning Optical Flow



Why?

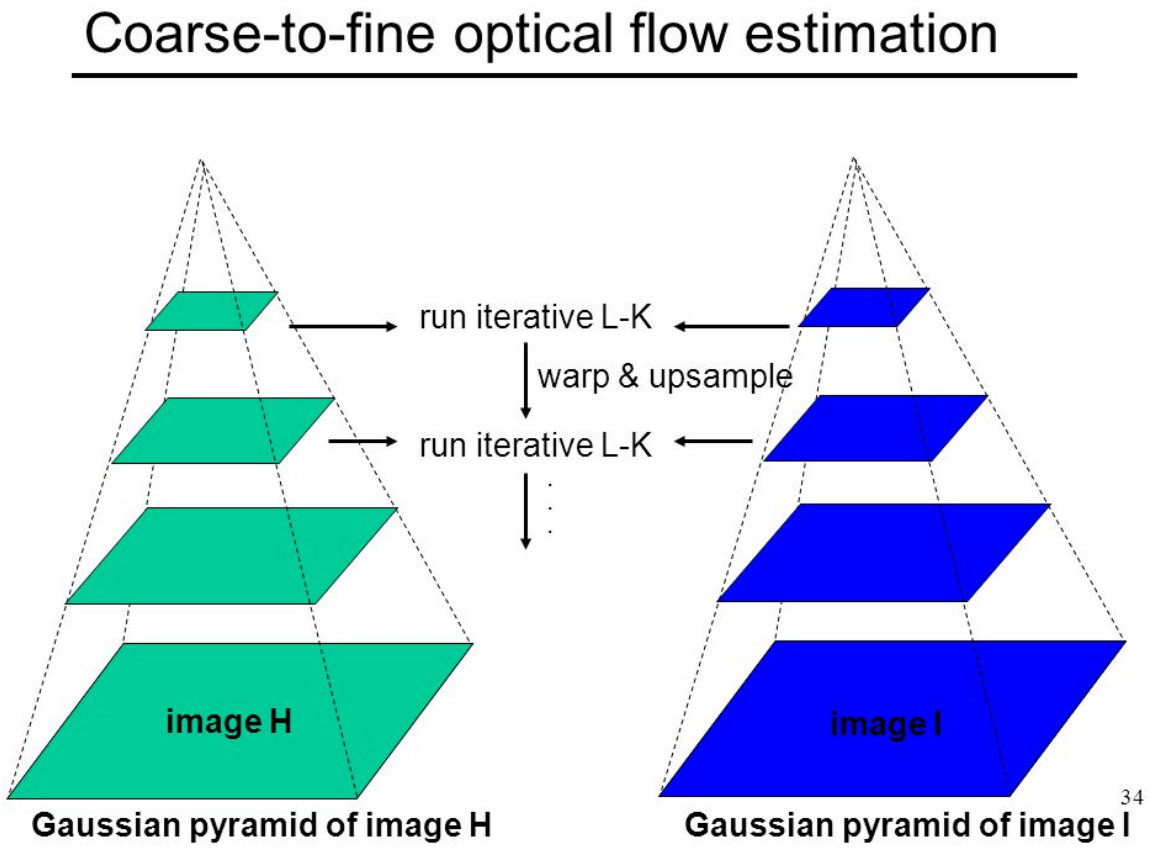
- Learn good features for matching
- Scene reasoning (e.g. occlusion, segmentation, semantics)

Large Motion

Standard CNNs only have small receptive field

Standard solution:
multi-resolution + warping

Multi-resolution + Warping



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

$$\tilde{J}(x) = J(x + F(x))$$

$$F(\mathbf{p}_1) = 0, \quad F(\mathbf{p}_2) = \mathbf{p}_1 - \mathbf{p}_2, \tag{1}$$

we have

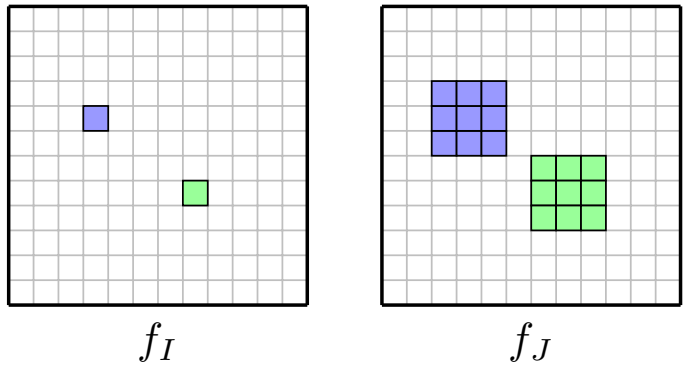
$$\begin{aligned} \tilde{J}(\mathbf{p}_1) &= J(\mathbf{p}_1 + F(\mathbf{p}_1)) && (2)\\ &= J(\mathbf{p}_1 + 0) = J(\mathbf{p}_1), && (3) \end{aligned}$$

$$\begin{aligned} \tilde{J}(\mathbf{p}_2) &= J(\mathbf{p}_2 + F(\mathbf{p}_2)) && (4)\\ &= J(\mathbf{p}_2 + \mathbf{p}_1 - \mathbf{p}_2) = J(\mathbf{p}_1). && (5) \end{aligned}$$

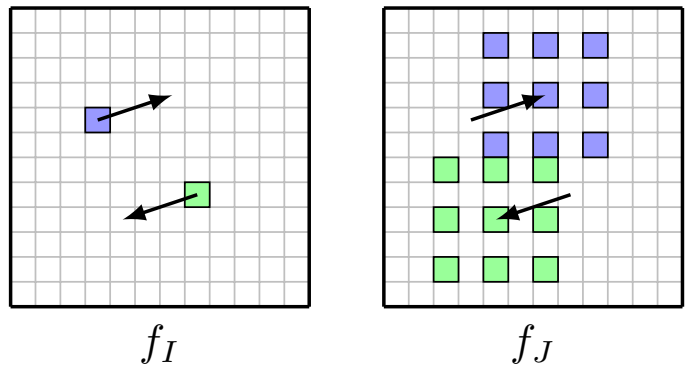
$$\tilde{J}(\mathbf{p}_1) = \tilde{J}(\mathbf{p}_2) = J(\mathbf{p}_1)$$

Solution

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

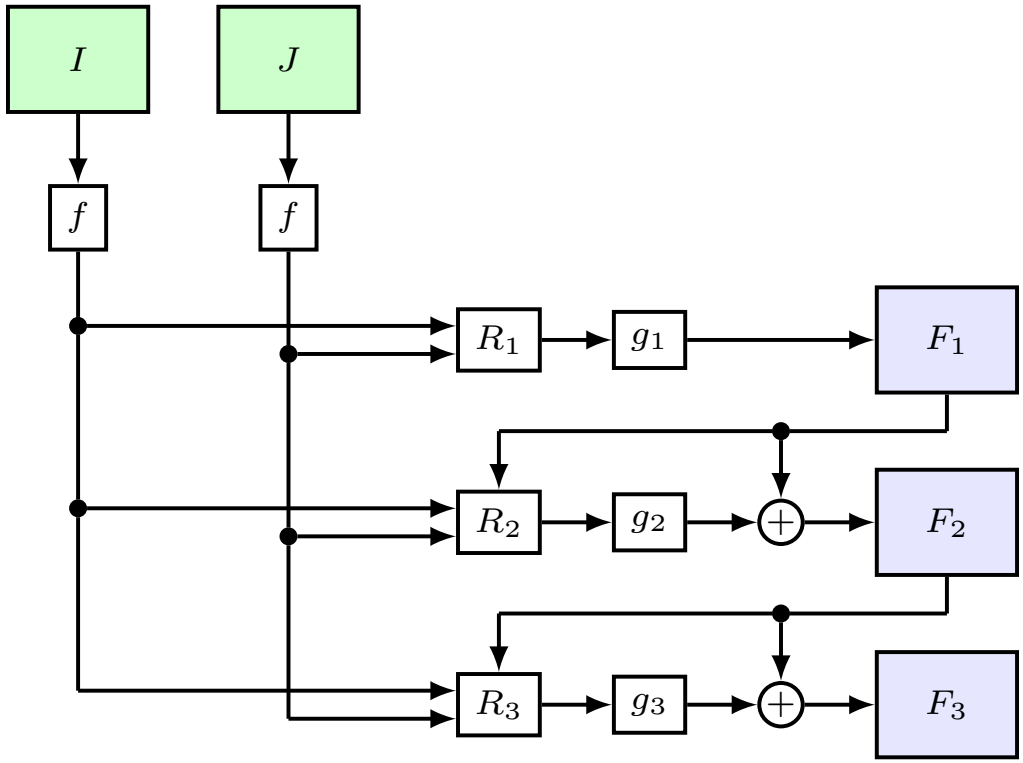


(e) Standard cost volume



(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