THE UNIVERSITY OF TEXAS AT DALLAS

Deep Optical Flow Networks and Applications

CS 6384 Computer Vision

Professor Yapeng Tian

Department of Computer Science

Optical Flow is the pattern of apparent motion of objects, surfaces, and edges in a visual scene caused by the relative motion between an observer and a scene

Lucas-Kanade Method for Optical Flow Estimation

- Brightness Constancy: the intensity or brightness of a pixel remains constant while moving from one frame to another
- Small Motion: the motion between consecutive frames is small
- Spatial Coherence: neighboring pixels have similar motion

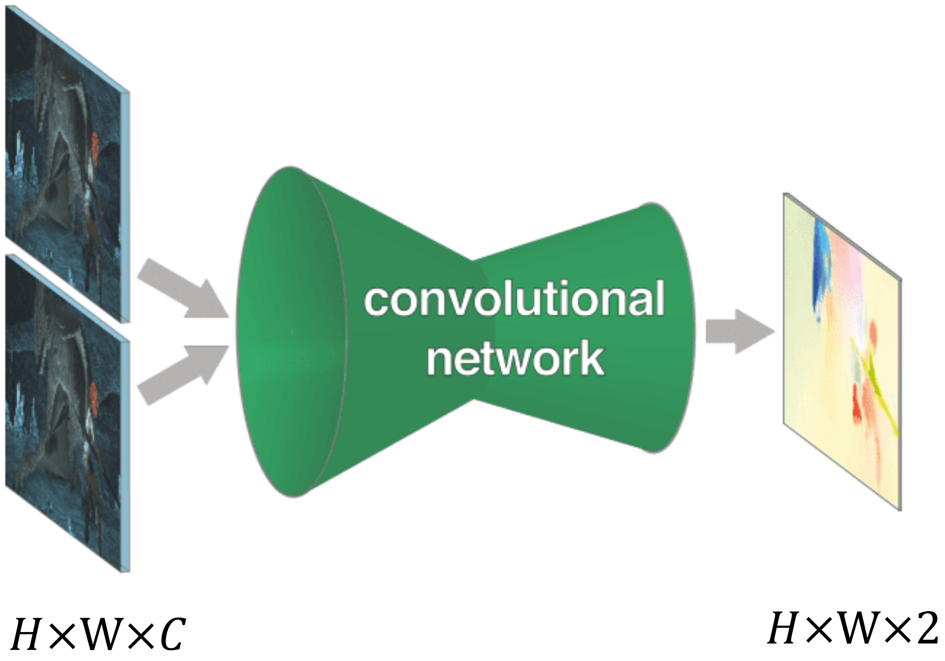
$$I_x u + I_y v + I_t = 0$$

$$\underset{\substack{A \\ 25\times2}}{\begin{bmatrix} I_x(\mathbf{p}_1) & I_y(\mathbf{p}_1) \\ I_x(\mathbf{p}_2) & I_y(\mathbf{p}_2) \\ \vdots & \vdots \\ I_x(\mathbf{p}_{25}) & I_y(\mathbf{p}_{25}) \end{bmatrix}} \underset{\substack{d \\ 2\times1}}{\begin{bmatrix} u \\ v \end{bmatrix}} = - \underset{\substack{b \\ 25\times1}}{\begin{bmatrix} I_t(\mathbf{p}_1) \\ I_t(\mathbf{p}_2) \\ \vdots \\ I_t(\mathbf{p}_{25}) \end{bmatrix}}$$

Can we use deep nets to estimate optical flow?

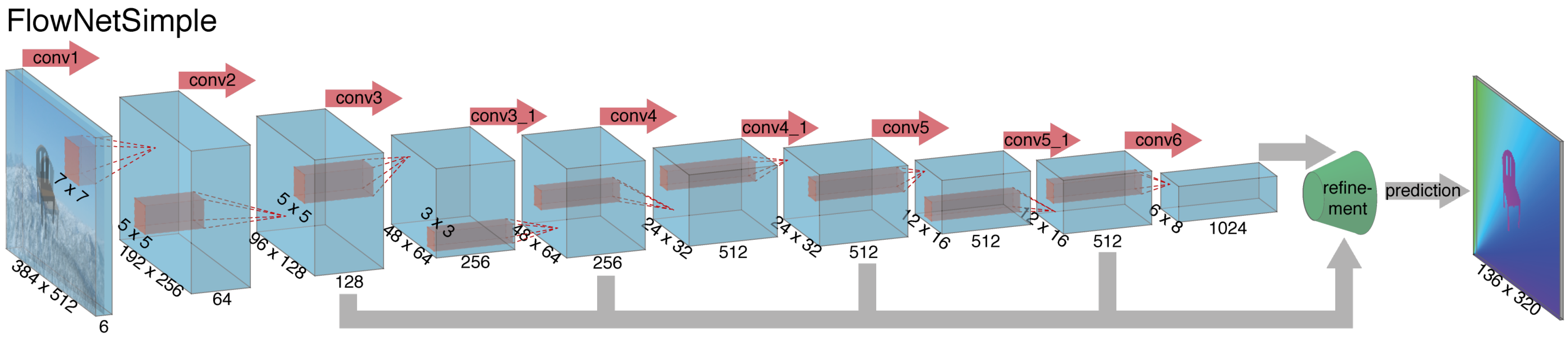
Estimating Optical Flow using Deep Networks

- Given two consecutive image frames: I_t and I_{t+1}, we aim to estimate the motion field (u, v) between them for each pixel



Regression Loss: L2, L1, …

FlowNet



Stack two images

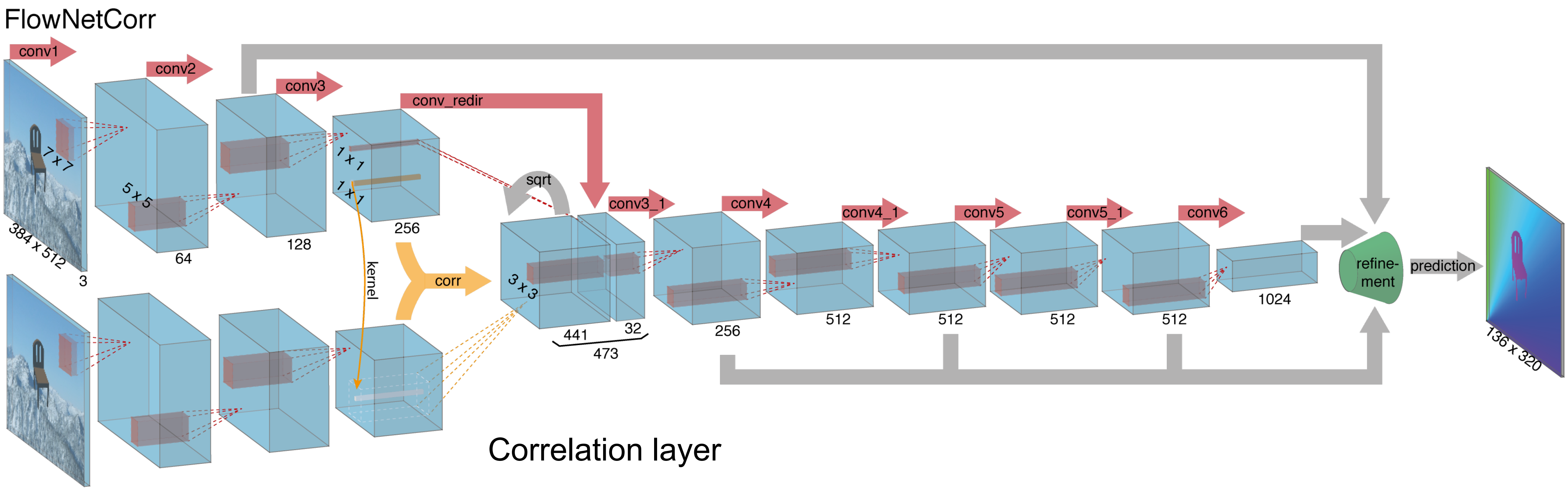
x-y flow fields

The architecture is similar to FCN for semantic segmentation

$$\frac{dx}{dt}, \frac{dy}{dt} = (\mathrm{u}, \mathrm{v})$$

FlowNet: Learning Optical Flow with Convolutional Networks. Fischer et al., ICCV, 2015

FlowNet



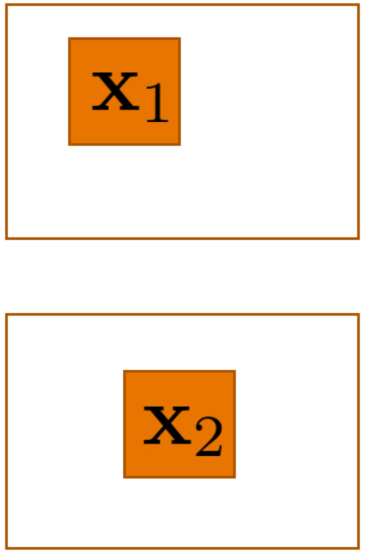
Correlation layer

FlowNet: Learning Optical Flow with Convolutional Networks. Fischer et al., ICCV, 2015

FlowNet

Correlation layer: multiplicative patch comparison between two feature maps

$$c(\mathbf{x}_1, \mathbf{x}_2) = \sum_{\mathbf{o} \in [-k,k] \times [-k,k]} \langle \mathbf{f}_1(\mathbf{x}_1 + \mathbf{o}), \mathbf{f}_2(\mathbf{x}_2 + \mathbf{o}) \rangle$$

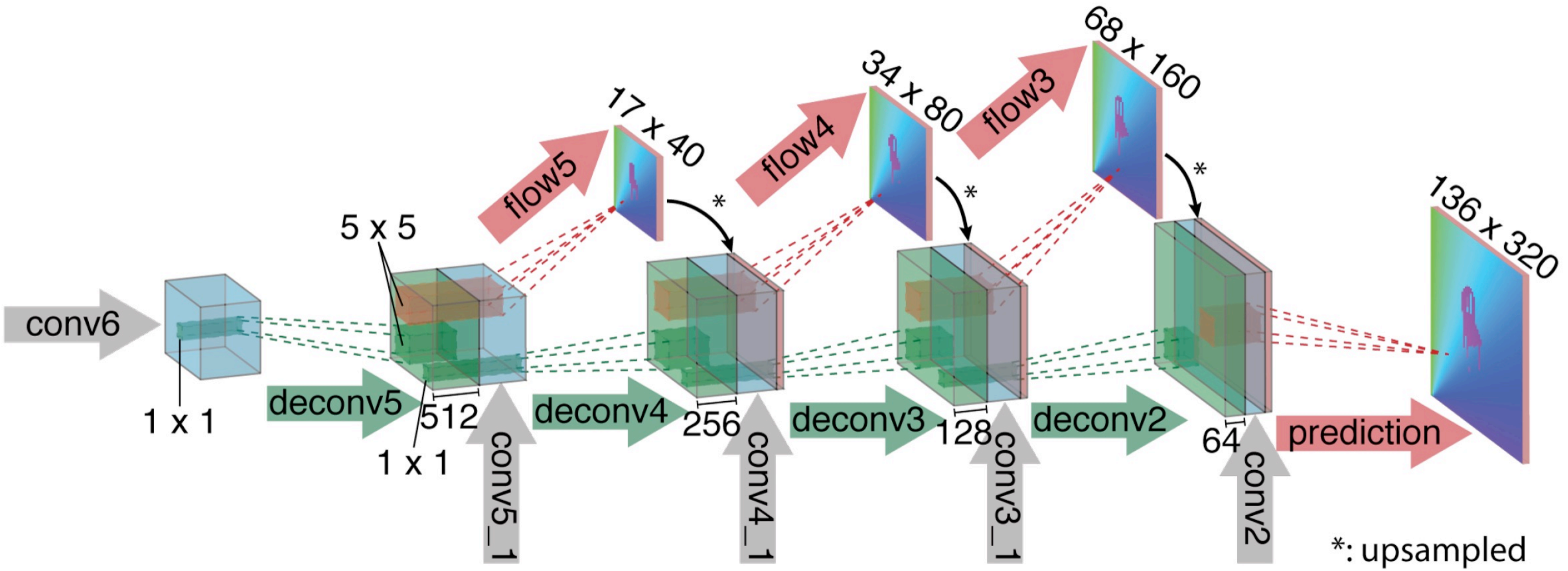


- Two patches centered at x1 and x2, with size K = 2k + 1
- Convolve data with another data
- Limit the patches for comparison with maximum displacement d
- Only compare patches in a neighborhood with size D = 2d + 1
- Output size $(w \times h \times D^2)$

FlowNet: Learning Optical Flow with Convolutional Networks. Fischer et al., ICCV, 2015

FlowNet

Refinement

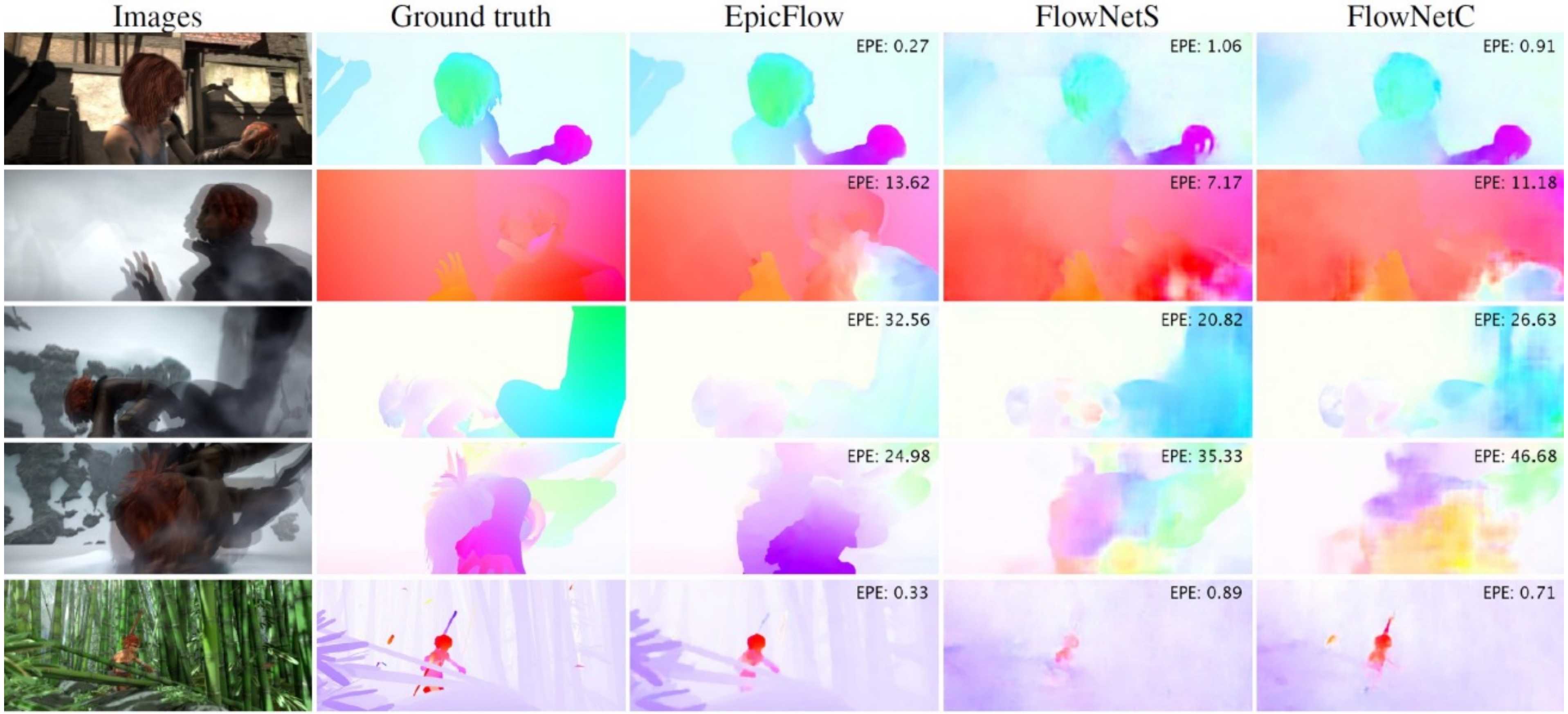


FlowNet: Learning Optical Flow with Convolutional Networks. Fischer et al., ICCV, 2015

Training Data

Flying Chairs Dataset

Deep Optical Flow Results



Results on Sintel (standard benchmark)

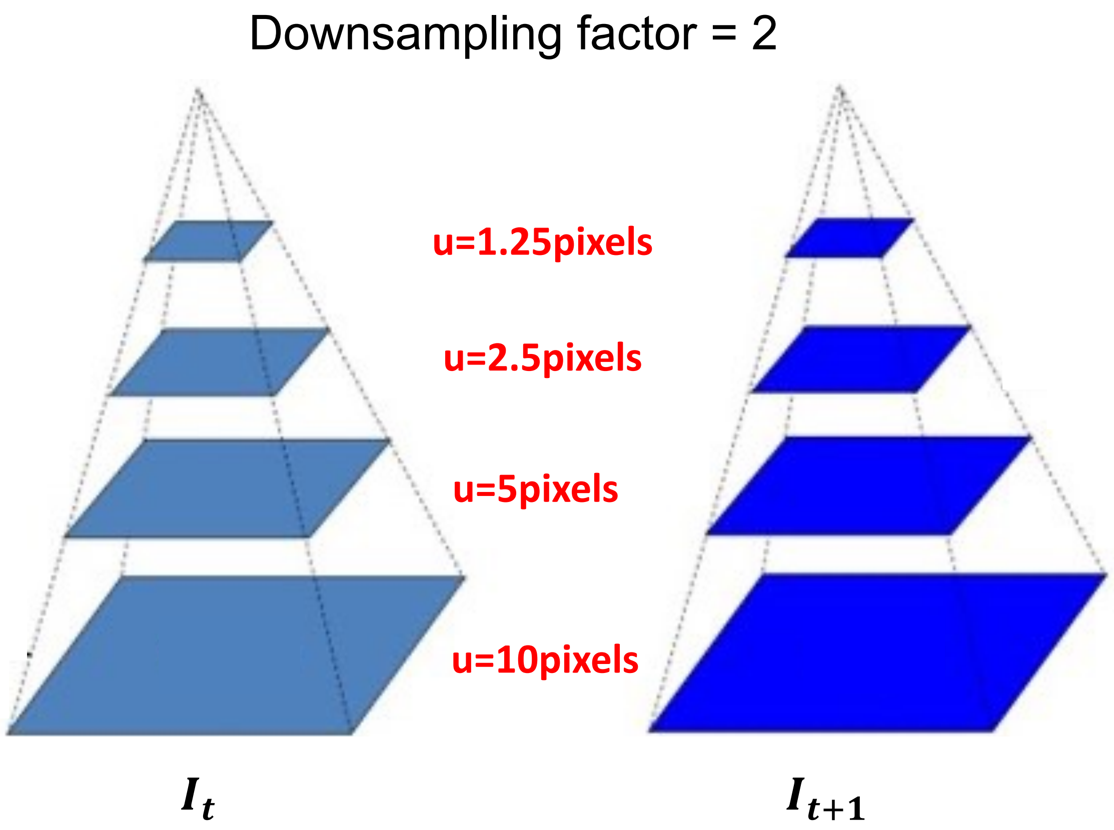
Revisiting the Small Motion Assumption

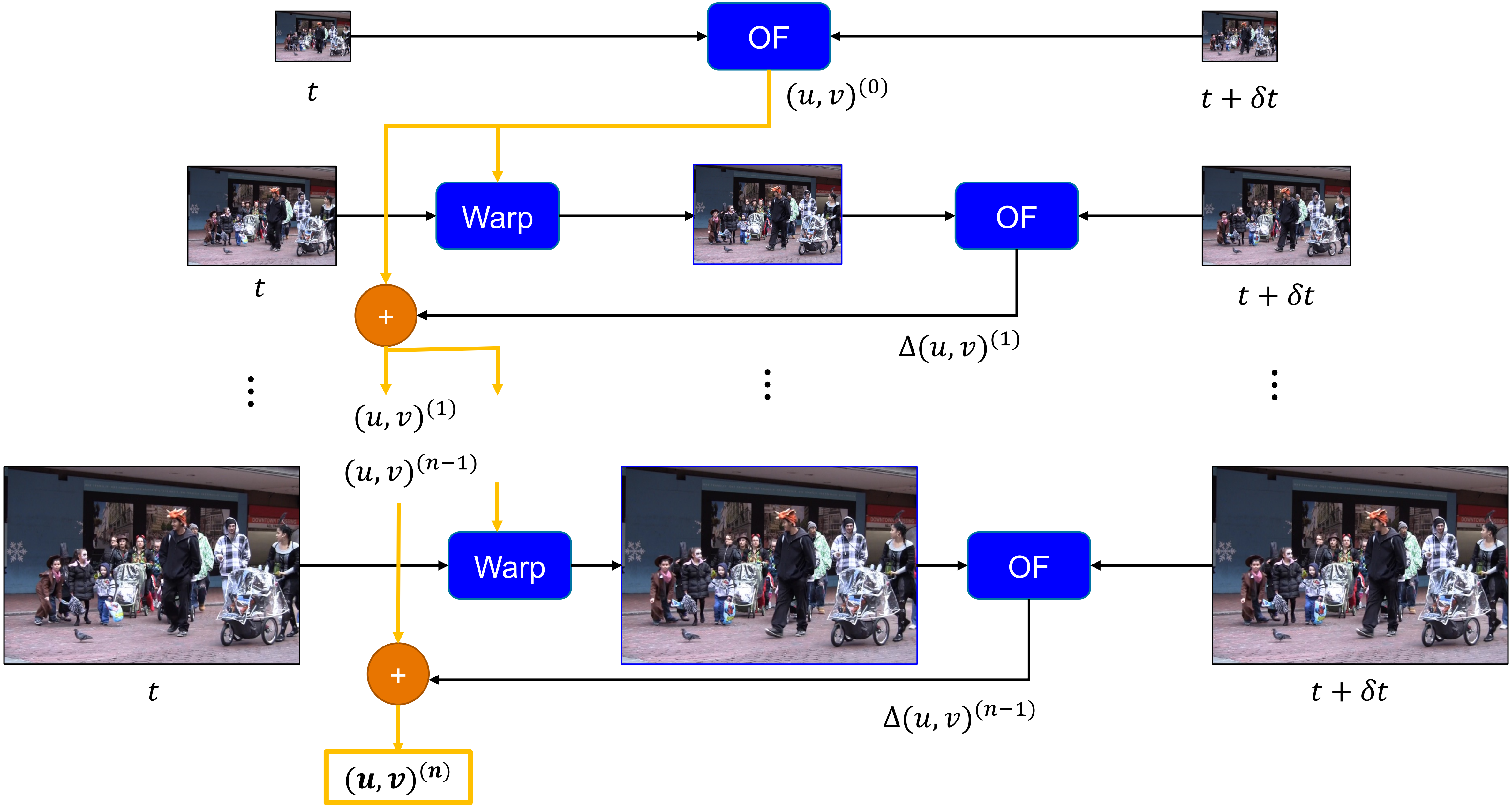
- Is this motion small enough?
  - Probably not—it's much larger than one pixel (2nd order terms dominate)
  - How to solve this problem?

$$I(x+\Delta x, y+\Delta y, t+\Delta t) = I(x,y,t) + \frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t + \text{higher-order terms}$$

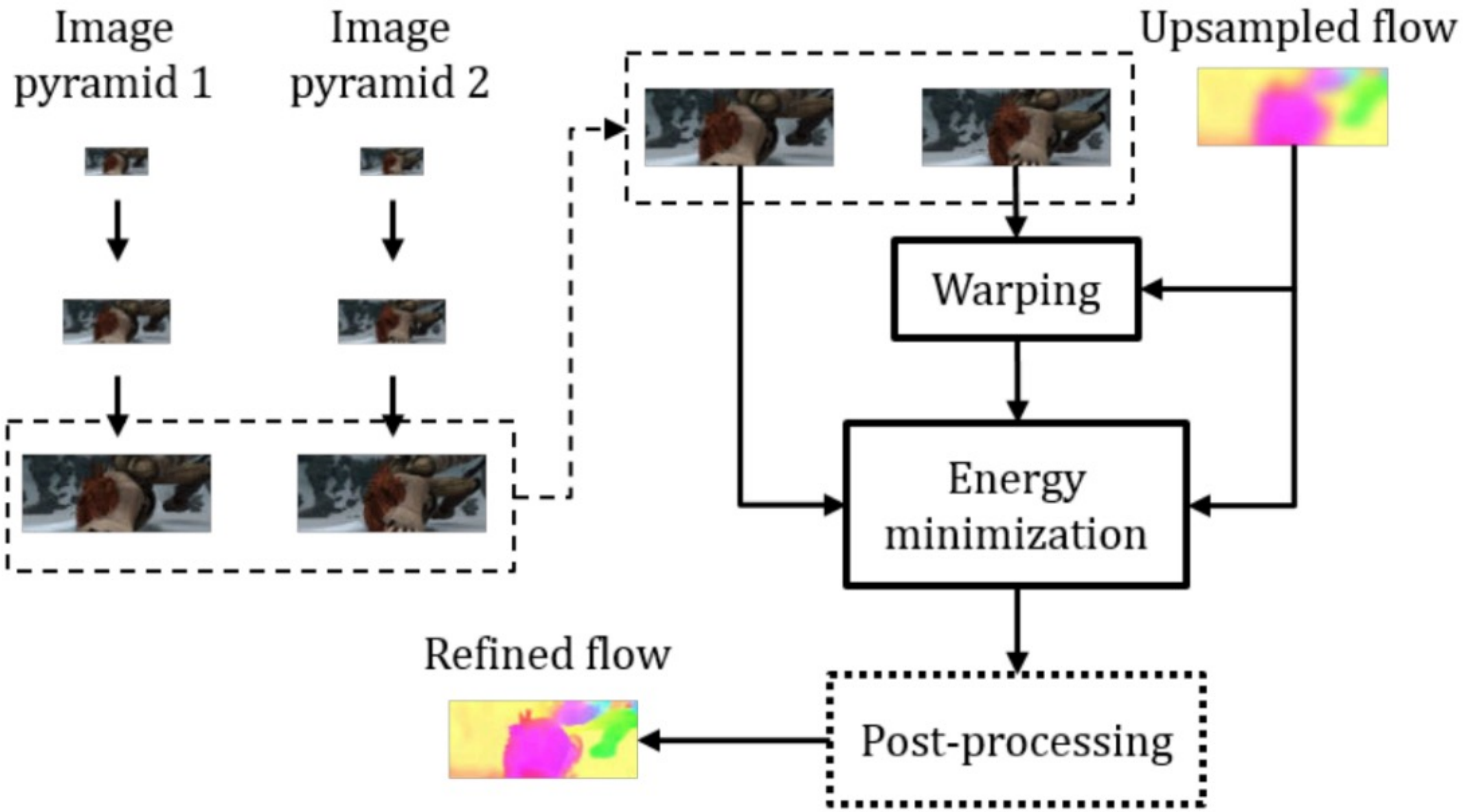
High-order terms will have large values for large motion

Coarse-to-fine Optical Flow Estimation



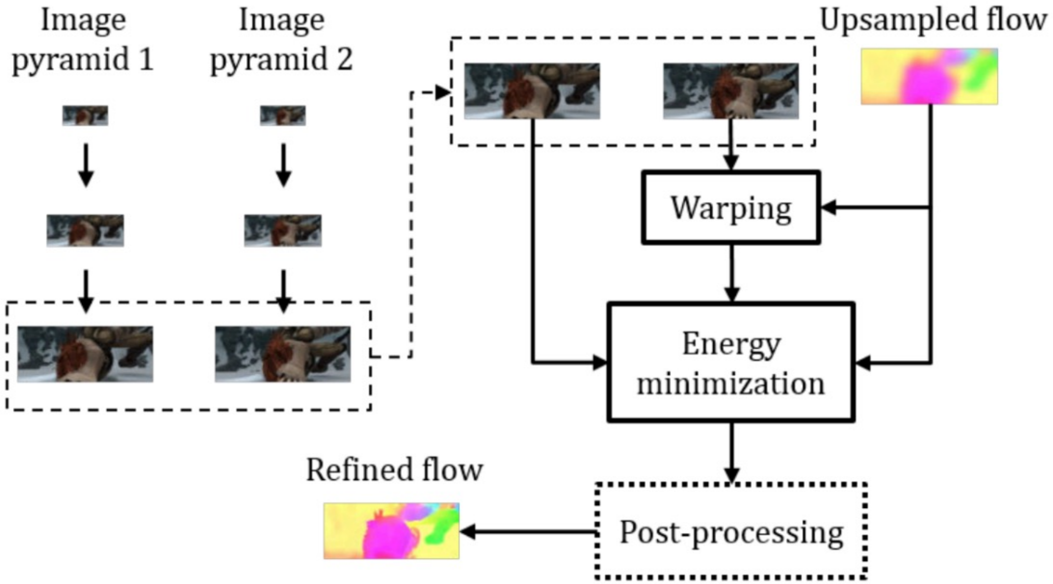

OF
t
$(u,v)^{(0)}$
$t+\delta t$
Warp
OF
t
+
$t+\delta t$
$\Delta(u,v)^{(1)}$
$(u,v)^{(1)}$
$(u,v)^{(n-1)}$
Warp
OF
+
t
$t+\delta t$
$\Delta(u,v)^{(n-1)}$
$(\boldsymbol{u},\boldsymbol{v})^{(\boldsymbol{n})}$

Coarse-to-fine Optical Flow Estimation

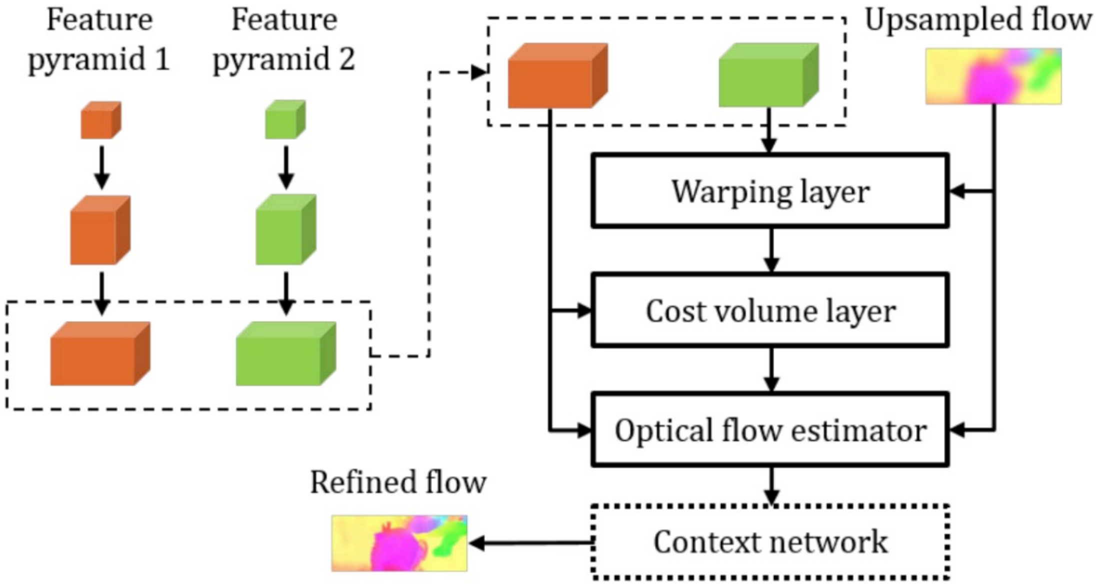


Simplified illustration

Coarse-to-fine Optical Flow Estimation



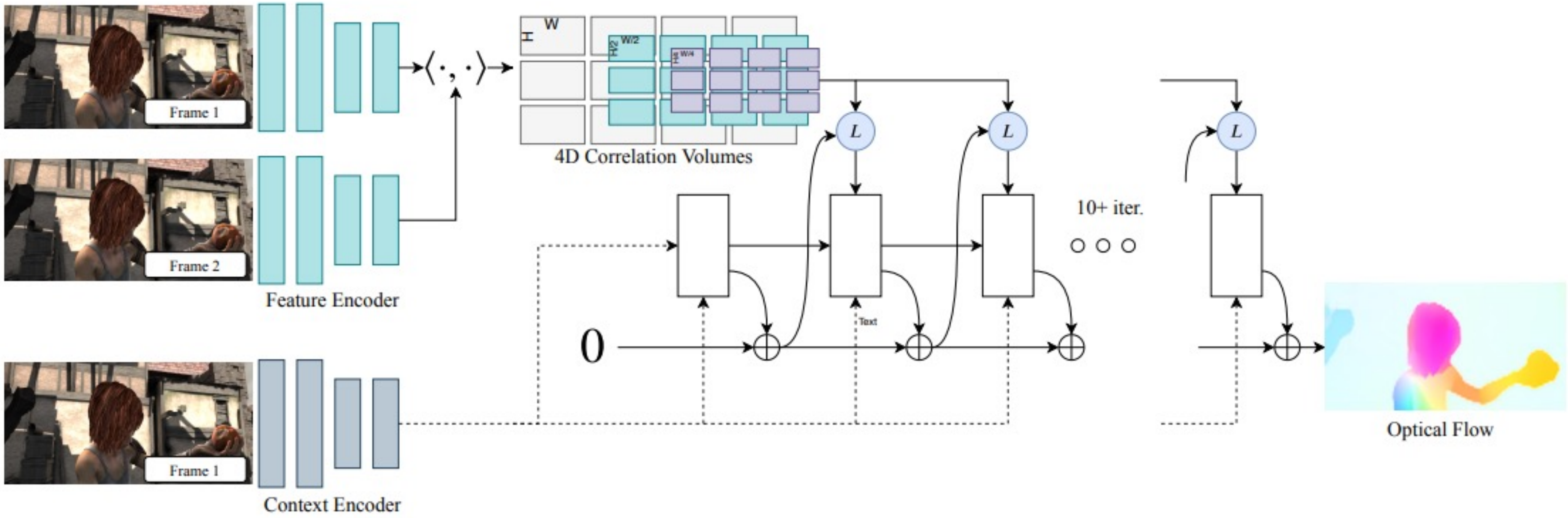
Traditional coarse-to-fine flow



PWC-net

[Sun et al., "PWC-Net", 2018]

RATF



Recurrent All-Pairs Field Transforms (RAFT), a new deep network architecture for optical flow

[Teed and Deng. “RATF”, 2020]

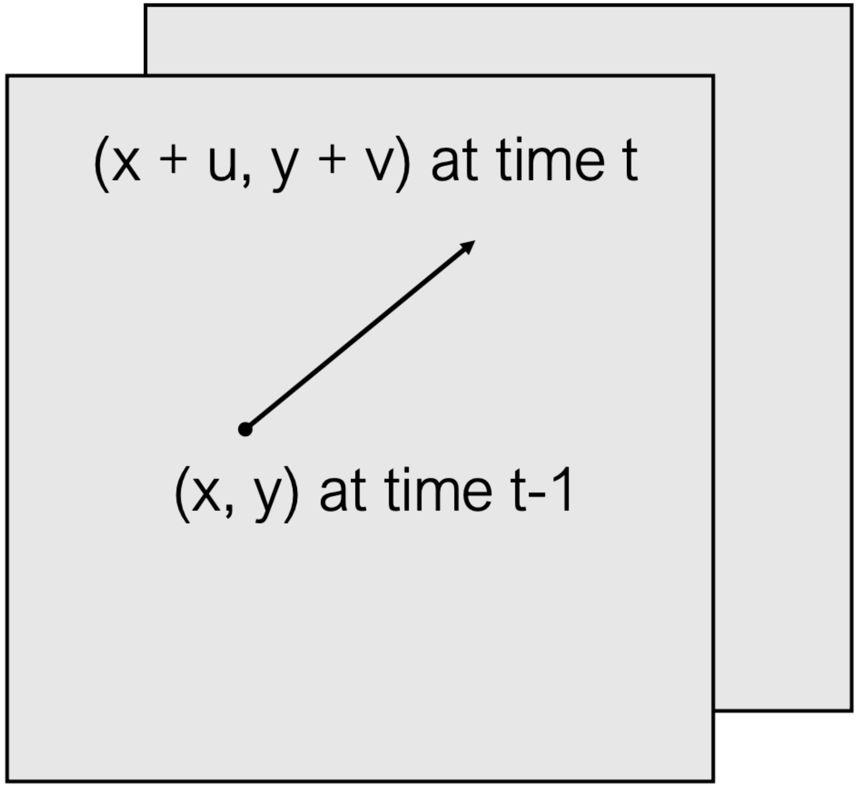
Applications

- Video Stabilization
- Video Frame Interpolation
- Action Recognition
- Video Restoration
- Visual Tracking
- …

Video Stabilization – Remove Camera Shake

https://cseweb.ucsd.edu/~ravir/jiyang_cvpr20.pdf [Yu and Ramamoorthi, 2020]

Video Frame Interpolation

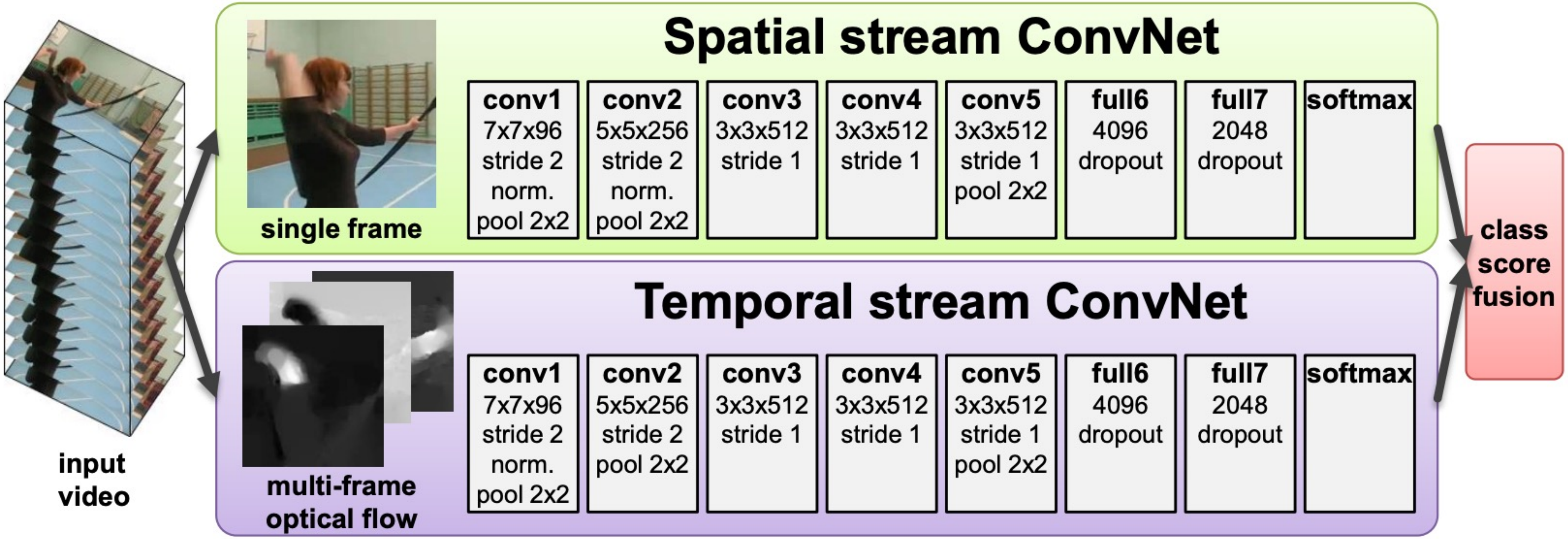


- use flow to estimate where pixel will be between two frames
- Synthesize intermediate frames to generate slow-motion videos

Credit: Shu Kong

https://www.youtube.com/watch?v=MjViy6kyiqs

Action Recognition

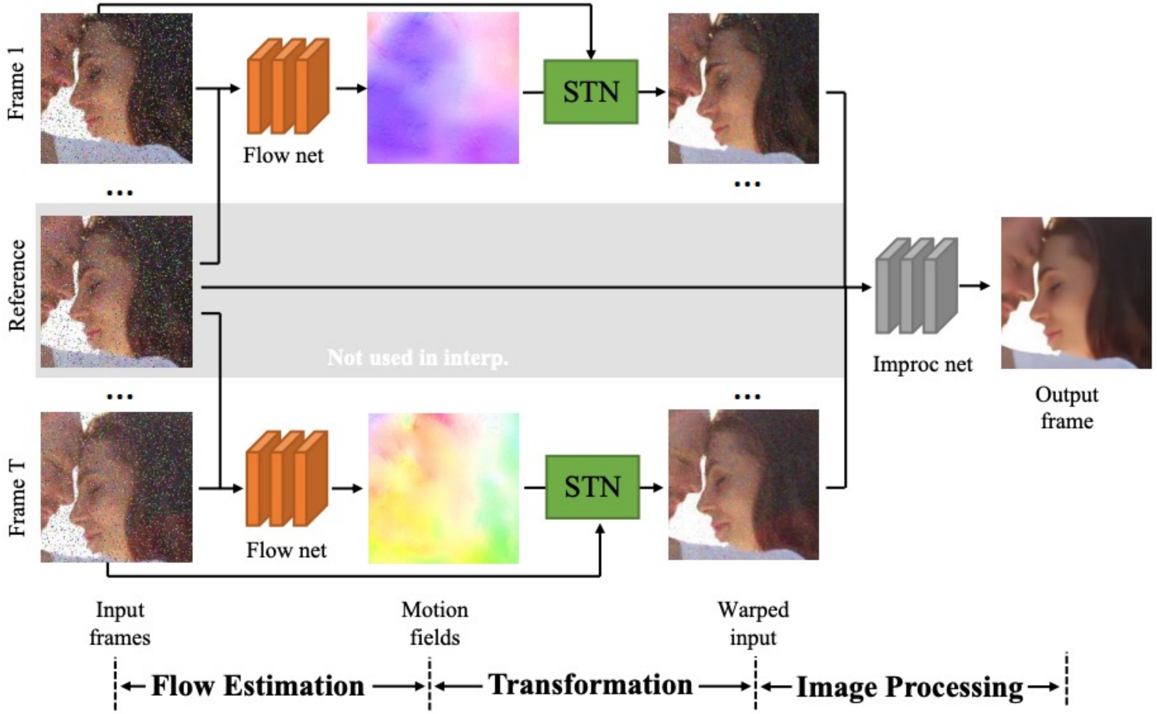


Two-stream architecture for video classification

[Simonyan and Zisserman, 2014]

Video Restoration

Optical flow can be used to address a series video restoration tasks, such as denoising, deblocking, and super-resolution



- Flow net to estimate motion field between neighboring frames
- Stack warped frames as input for the image processing network to predict the high-quality frame

Video Restoration

https://www.youtube.com/watch?v=msC5GK9aV9Q

Visual Tracking

https://nanonets.com/blog/optical-flow/

Further Reading

FlowNet: Learning Optical Flow with Convolutional Networks, 2015
https://arxiv.org/abs/1504.06852

PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume, 2018
https://arxiv.org/abs/1709.02371

RAFT: Recurrent All-Pairs Field Transforms for Optical Flow, 2020
https://arxiv.org/abs/2003.12039