Computer Vision; Image Transformation; Optical Flow and Depth Estimation

Maziar Raissi

Assistant Professor

Department of Applied Mathematics

University of Colorado Boulder

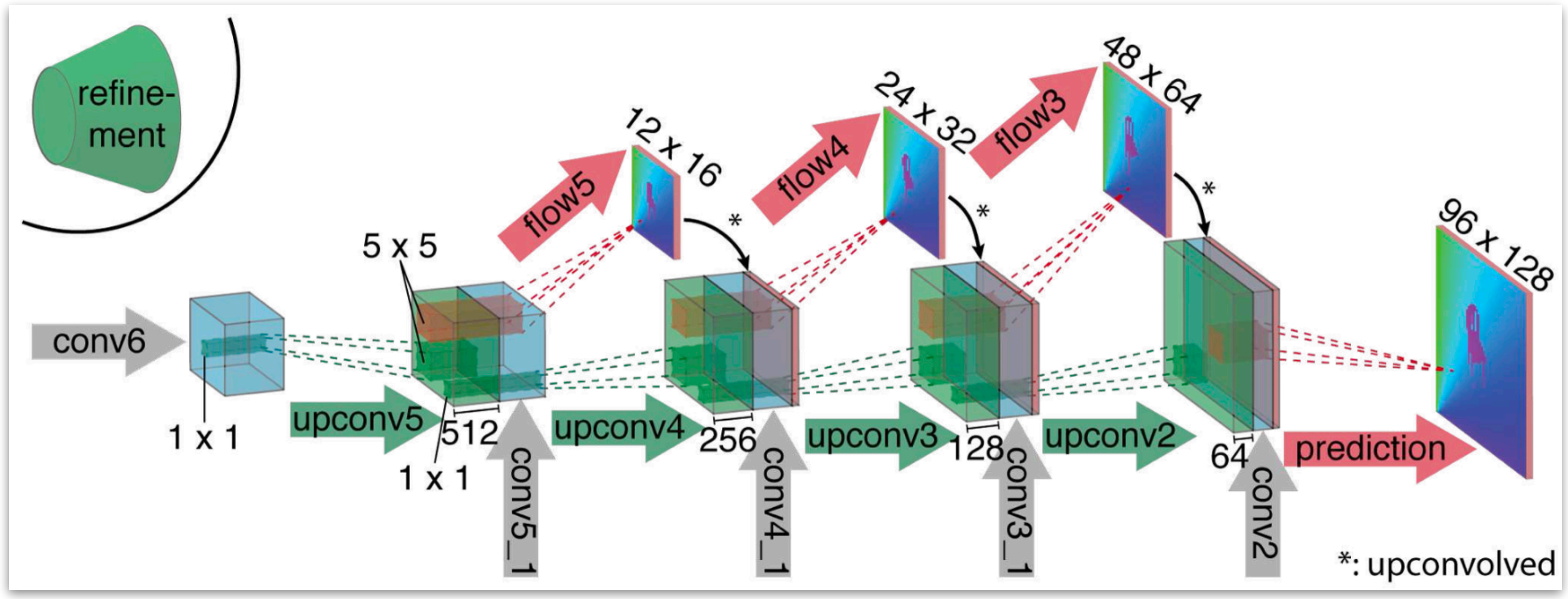
maziar.raissi@colorado.edu

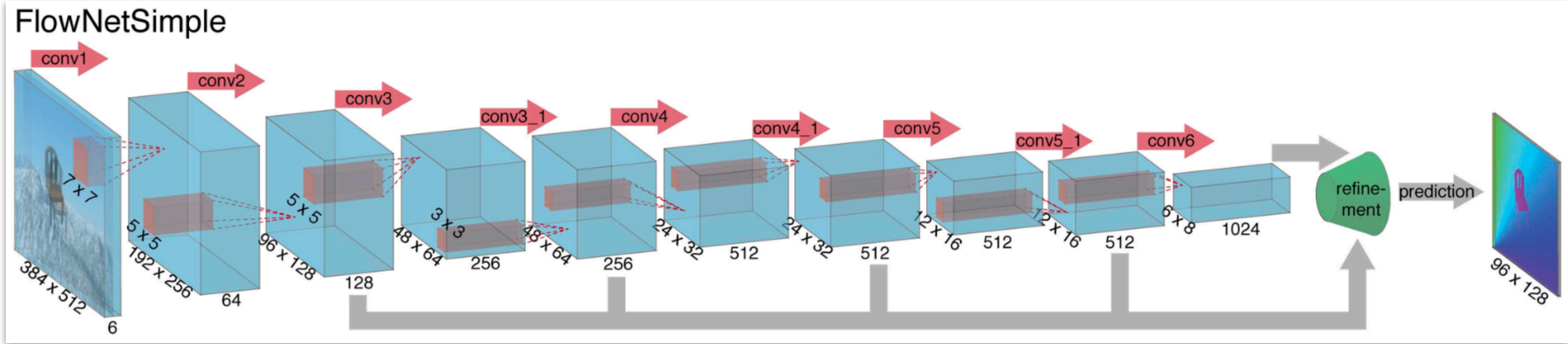
FlowNet: Learning Optical Flow with Convolutional Networks

Given a dataset consisting of image pairs and ground truth flows, train a network to predict the x-y flow fields directly from the images.

Flying Chairs Dataset | Data Augmentation

upconv: unpooling + conv





	Frame pairs	Frames with ground truth	Ground truth density per frame
Middlebury	72	8	100%
KITTI	194	194	∽50%
Sintel	1,041	1,041	100%
Flying Chairs	22,872	22,872	100%

Images | Ground truth | FlowNetS

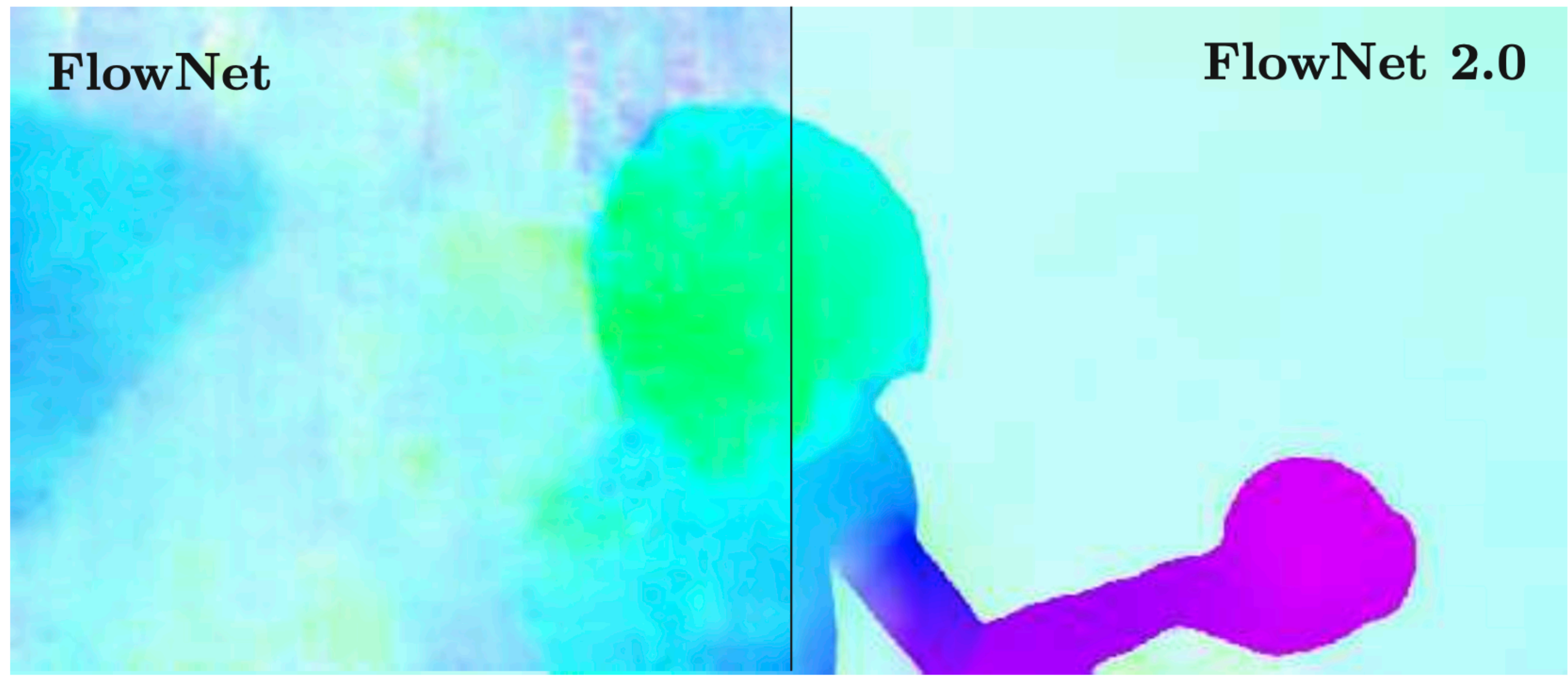
Method	Sintel Clean train	Sintel Clean test	Sintel Final train	Sintel Final test	KITTI train	KITTI test	Middlebury train AEE	Middlebury train AAE	Middlebury test AEE	Middlebury test AAE	Chairs test	Time (sec) CPU	Time (sec) GPU
EpicFlow [30]	2.27	4.12	3.57	6.29	3.47	3.8	0.31	3.24	0.39	3.55	2.94	16	-
DeepFlow [35]	3.19	5.38	4.40	7.21	4.58	5.8	0.21	3.04	0.42	4.22	3.53	17	-
EPPM [3]	-	6.49	-	8.38	-	9.2	-	-	0.33	3.36	-	-	0.2
LDOF [6]	4.19	7.56	6.28	9.12	13.73	12.4	0.45	4.97	0.56	4.55	3.47	65	2.5
FlowNetS	4.50	7.42	5.45	8.43	8.26	-	1.09	13.28	-	-	2.71	-	0.08
FlowNetS+v	3.66	6.45	4.76	7.67	6.50	-	0.33	3.87	-	-	2.86	-	1.05
FlowNetS+ft	(3.66)	6.96	(4.44)	7.76	7.52	9.1	0.98	15.20	-	-	3.04	-	0.08
FlowNetS+ft+v	(2.97)	6.16	(4.07)	7.22	6.07	7.6	0.32	3.84	0.47	4.58	3.03	-	1.05

End Point Error (EPE) Loss: Euclidean distance between the predicted flow vector and the ground truth, averaged over all pixels

Dosovitskiy, Alexey, et al. "Flownet: Learning optical flow with convolutional networks." *Proceedings of the IEEE international conference on computer vision*. 2015.

Boulder

FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks



– problems with small displacements

– noisy artifacts in estimated flow fields

Dataset Schedules

Numbers indicate endpoint errors on Sintel train clean

Architecture	Datasets	S_{short}	S_{long}	S_{fine}
FlowNetS	Chairs	4.45	-	-
	Chairs	-	4.24	4.21
	Things3D	-	5.07	4.50
	mixed	-	4.52	4.10
	Chairs→Things3D	-	4.24	**3.79**
FlowNetC	Chairs	3.77	-	-
	Chairs→Things3D	-	3.58	**3.04**

```
Training on Chairs first and fine-tuning on Things3D yields
the best results. FlowNetC performs better than FlowNetS.
```

FlowNetCorr

$f_1, f_2 \in \mathbb{R}^{w \times h \times c}$ or $f_1, f_2 : \mathbb{R}^2 \to \mathbb{R}^c$

$c(x_1, x_2) \to$ correlation of two patches centered at x_1 in the first map and x_2 in the second map

$c(x_1, x_2) = \sum_{o \in [-k,k] \times [-k,k]} \langle f_1(x_1 + o), f_2(x_2 + o) \rangle \to$ for a square patch of size $K = 2k + 1$

$w^2 h^2 c K^2 \to$ computational cost

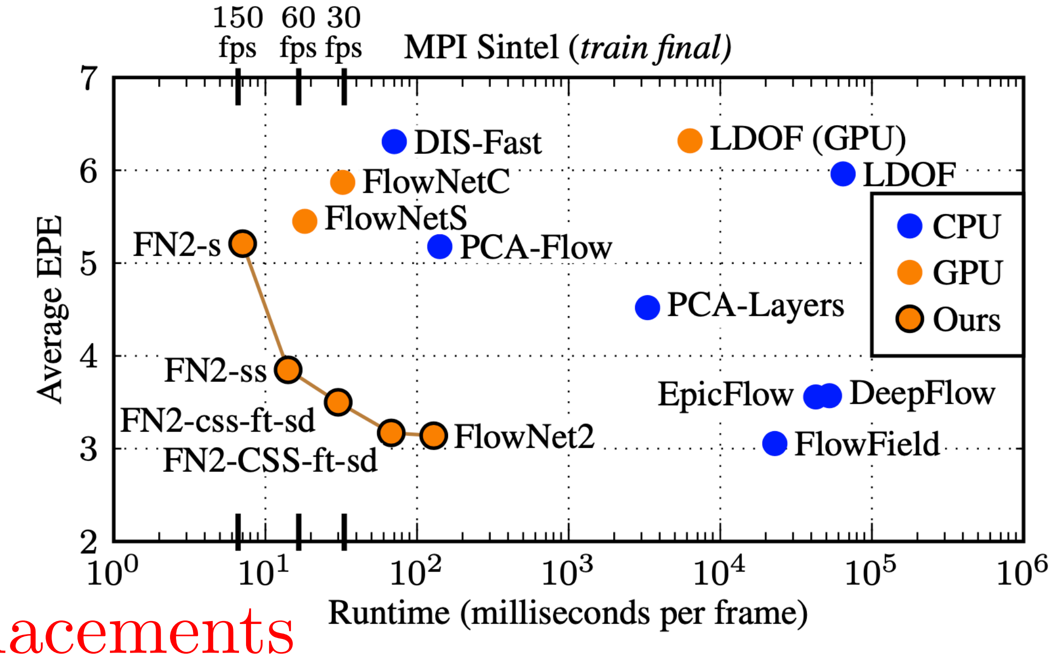
compute $c(x_1, x_2)$ only in a neighborhood of x_1 of size $D = 2d + 1$

$whD^2 cK^2 \to$ computational cost

$w \times h \times D^2 \to$ output size

use stride s_1 to quantize x_1 globally

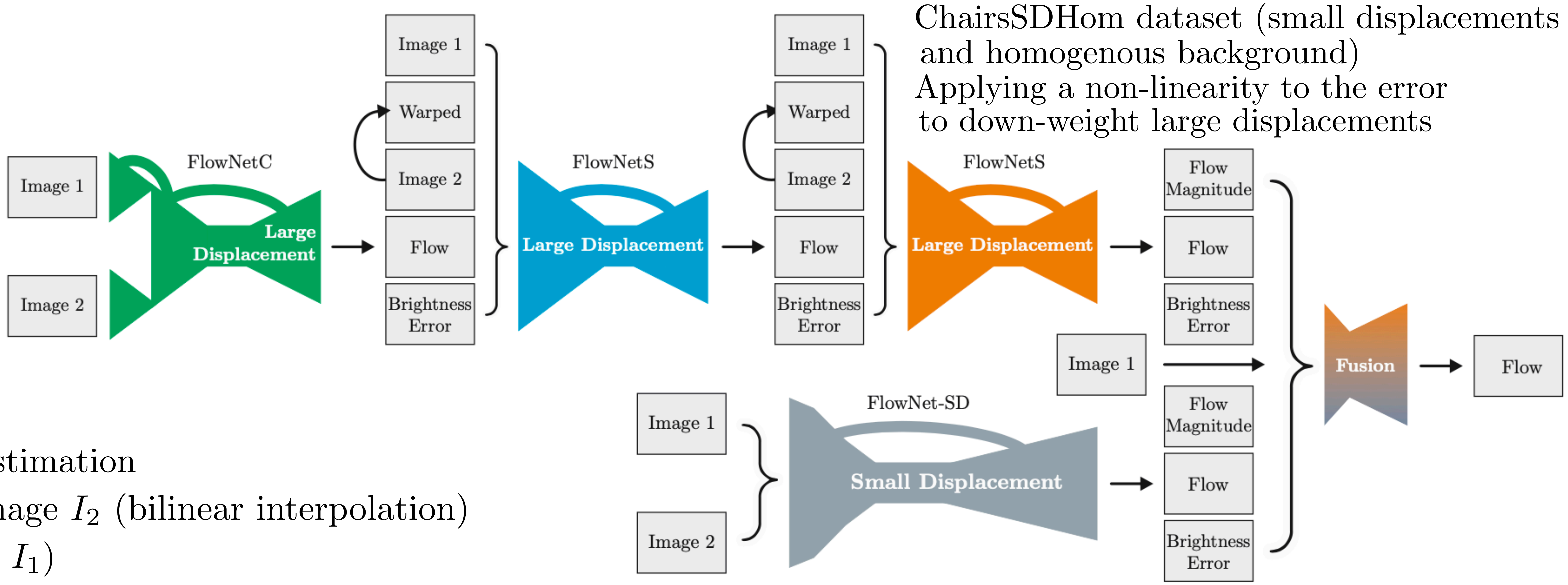
use stride s_2 to quantize x_2 locally around x_1



Small Displacements

ChairsSDHom dataset (small displacements and homogenous background)

Applying a non-linearity to the error to down-weight large displacements



Stacking Networks

$I_1, I_2 \to$ images $\quad w = (u, v)^T \to$ previous flow estimation

$\widetilde{I}_2(x, y) = I_2(x + u, y + v) \to$ warp the second image I_2 (bilinear interpolation)

$e = \|\widetilde{I}_2(x, y) - I_1\| \to$ error ($\widetilde{I}_2$ should be close to I_1)

Ilg, Eddy, et al. "Flownet 2.0: Evolution of optical flow estimation with deep networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

Questions?