LiteFlowNet: A Lightweight Convolutional Neural Network for Optical Flow Estimation

Tak-Wai Hui, Xiaoou Tang, Chen Change Loy
CUHK-SenseTime Joint Lab, The Chinese University of Hong Kong

SALT LAKE CITY • JUNE 18-22 2018 CVPR

1. Introduction

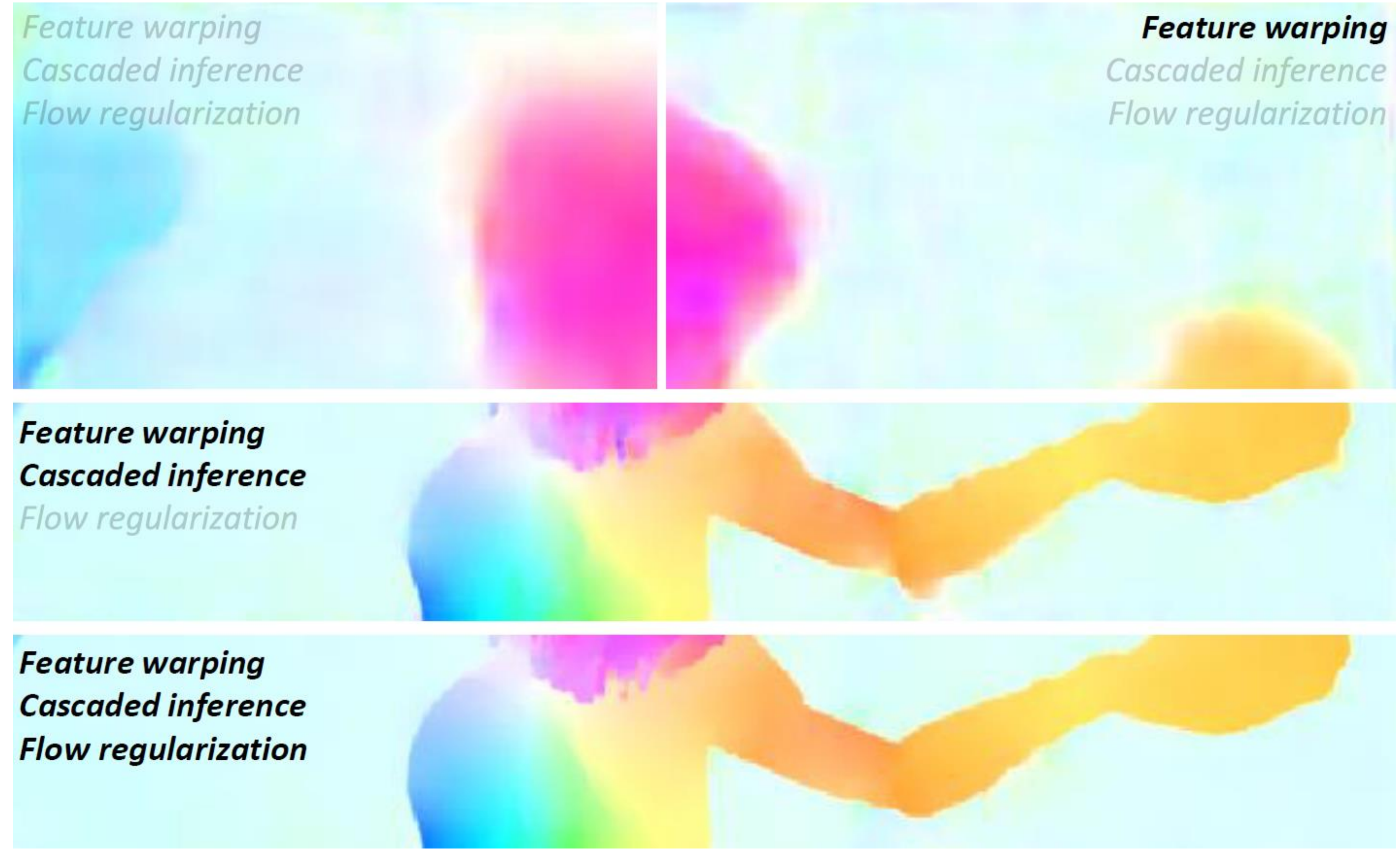
FlowNet2 (CVPR17)

1. Large network cascade (~160M)
2. U-Net architecture
3. Image warping per cascade
4. Feature matching per cascade (except FlowNetS)

LiteFlowNet (CVPR18)

1. Lightweight (~5M parameters)
2. Specialized architecture:
   - Data fidelity & regularization as variational methods
3. Feature warping per pyramid level
4. Cascaded flow inference per level
   - Descriptor matching & sub-pixel refinement
5. Flow regularization per level

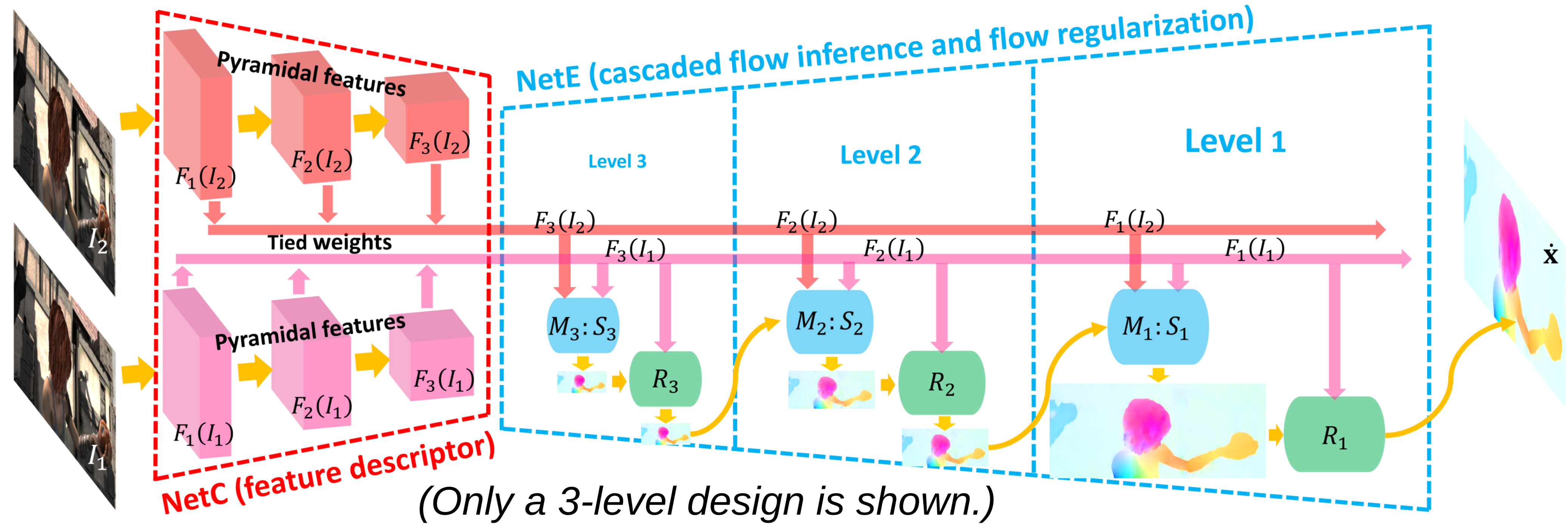
- LiteFlowNet outperforms FlowNet2 on the challenging Sintel final-pass and KITTI benchmarks.
- LiteFlowNet is ***30.26 times smaller in the model size*** and ***1.36 times faster in the running speed*** than FlowNet2.
- Our network innovates the useful elements from conventional methods:
  - Brightness constraint in data fidelity to **pyramidal feature extraction**.
  - Image warping to **feature warping**.
- We introduce a **cascaded flow inference** with **feature warping** and a **flow regularization** in each pyramid level.



Examples demonstrate the effectiveness of the proposed components.

Project page (paper, supp. material, demo video, and code):

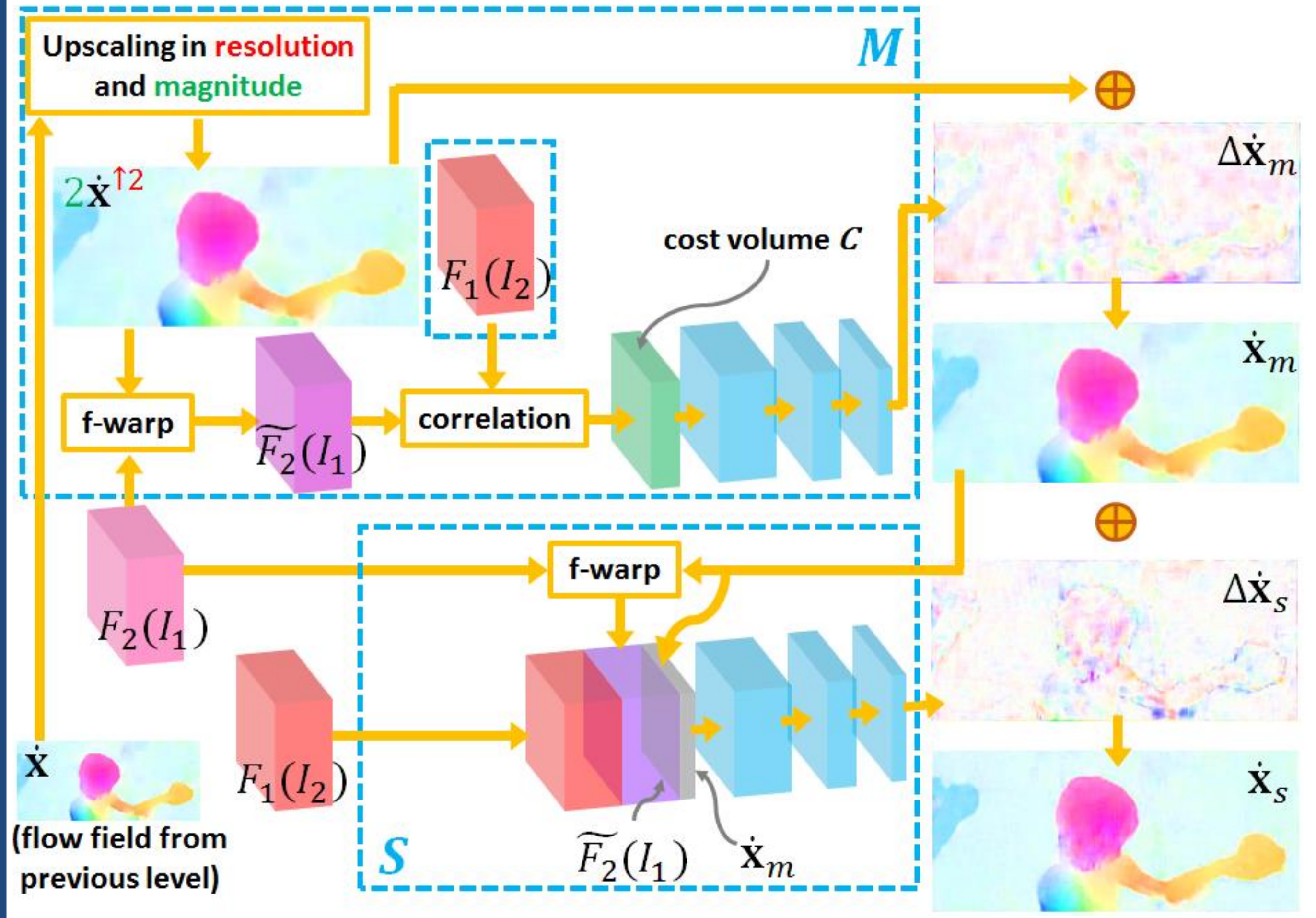
2. Network Details



(Only a 3-level design is shown.)

- NetC generates two pyramids of multi-scale high-level features.
- NetE yields multi-scale flow fields that each of them is generated by a **cascaded flow inference module** (**M**: descriptor matching + **S**: sub-pixel refinement) and a **regularization module R.**

2.1. Cascaded Flow Inference



i. **Descriptor Matching (M)** – Coarse flow inference by measuring feature correlation.

ii. **Sub-pixel Refinement (S)** – Refine previous pixel-level flow field.

Feature warping (**f-warp**) layer spatially displaces feature map F_2 towards F_1 according to the input flow field.

2.2. Flow Regularization

- Desired: A network can smooth flow field and maintain crisp flow boundaries as regularization term in conventional variational methods.
- We introduce a **feature-driven local convolution** (**f-lcon**) layer.
- A **feature-driven distance metric** $\mathcal{D}$ is trained to measure local flow variation from pyramidal feature, flow, and occlusion probability map.
- Channel c of a flow patch f is regularized by an adaptive **f-lcon** filter g:

$$f_g(x,y,c) = g(x,y,c) * f(x,y,c)\ ,$$

where

$$g(x,y,c) = \frac{\exp(-\mathcal{D}(x,y,c)^2)}{\sum_{(x_i,y_i)\in N(x,y)} \exp(-\mathcal{D}(x_i,y_i,c)^2)}.$$

3. Experimental Results

Method	Sintel clean		Sintel final		KITTI12		KITTI15			Middlebury	
	train	test	train	test	train	test	train	train (Fl-all)	test (Fl-all)	train	test
FlowNetS	4.50	7.42	5.45	8.43	8.26	-	-	-	-	1.09	-
FlowNetS-ft	(3.66)	6.96	(4.44)	7.76	7.52	9.1	-	-	-	0.98	-
FlowNetC	4.31	7.28	5.87	8.81	9.35	-	-	-	-	1.15	-
FlowNetC-ft	(3.78)	6.85	(5.28)	8.51	8.79	-	-	-	-	0.93	-
FlowNet2-S	3.79	-	4.99	-	7.26	-	14.28	51.06%	-	1.04	-
FlowNet2-C	3.04	-	4.60	-	5.79	-	11.49	44.09%	-	0.98	-
FlowNet2	**2.02**	**3.96**	**3.54**	6.02	4.01	-	10.08	29.99%	-	**0.35**	**0.52**
FlowNet2-ft-sintel	(1.45)	4.16	(2.19)	**5.74**	**3.54**	-	**9.94**	**28.02%**	-	**0.35**	-
FlowNet2-ft-kitti	3.43	-	4.83	-	(1.43)	**1.8**	(2.36)	(8.88%)	**11.48%**	0.56	-
SPyNet	4.12	6.69	5.57	8.43	9.12	-	-	-	-	0.33	0.58
SPyNet-ft	(3.17)	6.64	(4.32)	8.36	*3.36*	4.1	-	-	35.07%	0.33	0.58
LiteFlowNetX	3.58	-	4.79	-	6.38	-	15.81	34.90%	-	0.46	-
LiteFlowNet	**2.48**	-	**4.04**	-	**4.00**	-	**10.39**	**28.50%**	-	0.39	-
LiteFlowNet-ft	(1.35)	**4.54**	(1.78)	**5.38**	(1.05)	**1.6**	(1.62)	(5.58%)	**9.38%**	**0.30**	**0.40**