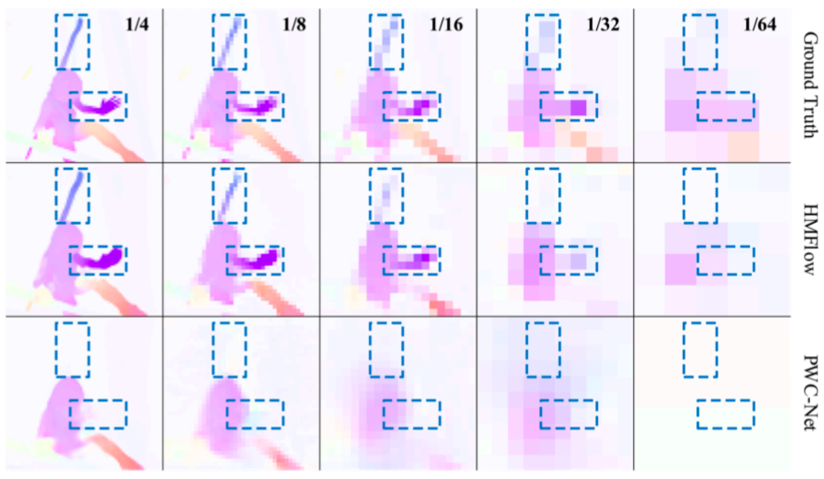
HMFlow: Hybrid Matching Optical Flow Network for Small and Fast-Moving Objects

Suihanjin Yu[1], Youmin Zhang[1], Chen Wang[1], Xiao Bai[1], Liang Zhang[1], Edwin R. Hancock[2]
[1]Beihang University, [2]University of York

Introduction



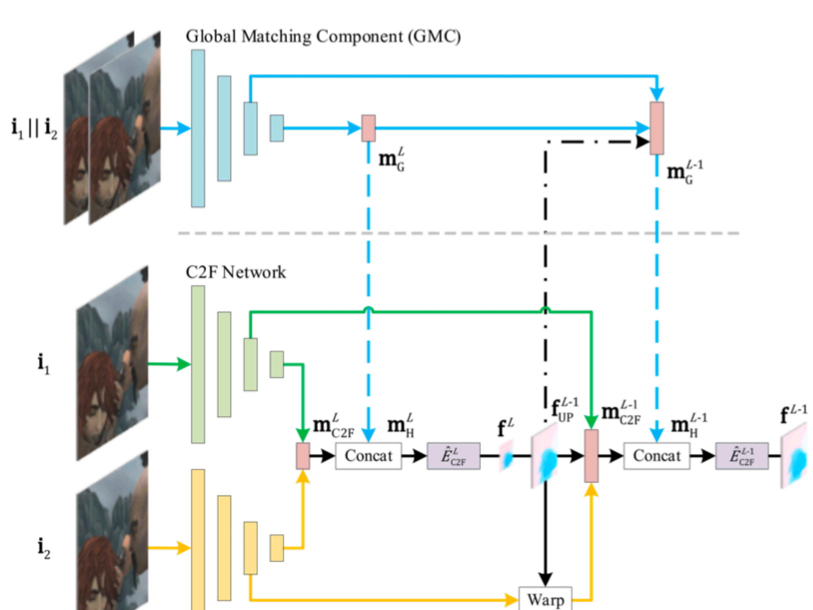
The Flow Spatial Pyramid

- The encoder-decoder based optical flow possess strong flexibility with large size of model parameters, causing high computing cost.
- Recently, coarse-to-fine architecture networks achieve high performances with relatively small model sizes, while they also inherit the problem of capturing the small and fast-moving objects from traditional coarse-to-fine methods.

To address these issues, we in this paper

- ✓ propose a lightweight but effective Global Matching Component (GMC) to produce global matching features that can cover the motion range of small and fast-moving objects;
- ✓ propose a Hybrid Matching Optical Flow Network (HMFlow), which integrates GMC to capture the small and fast-moving object;
- ✓ build a specific dataset, SFChairs, for flow estimation evaluation, especially for small and fast-moving object regions.

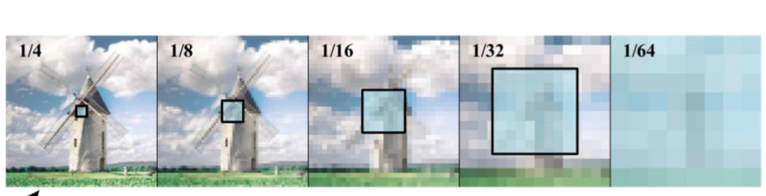
Architecture of HMFlow:



Key Component:

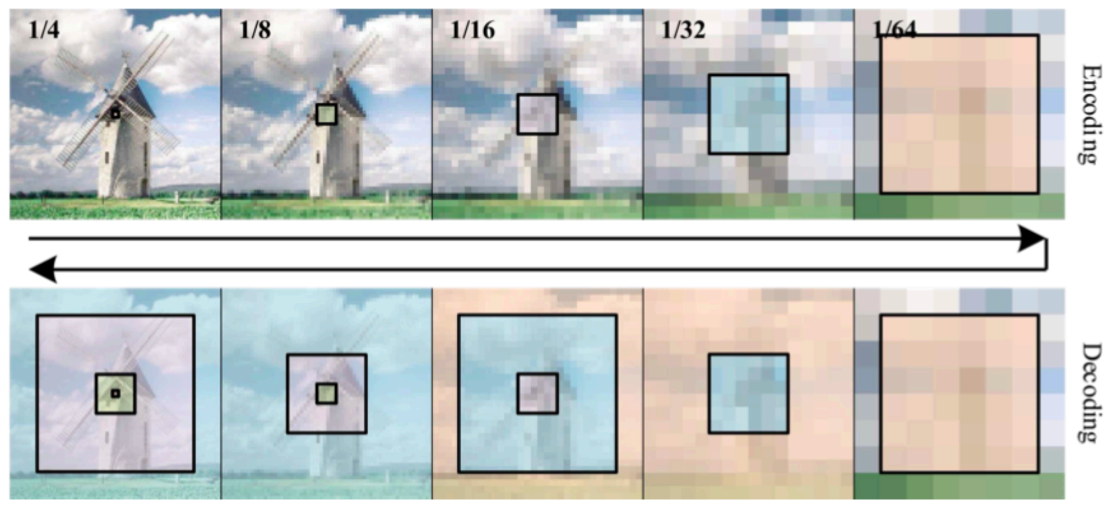
1. Feature fusion: $m_H^l = m_G^l \parallel m_{C2F}^l$
2. Flow estimation: $f^l = \hat{E}_{C2F}^l(m_H^l, u_{C2F}^l)$
3. Up-sampled flows f_{UP}^{L-1} are sent to m_G^{L-1} of GMC

The Local Search Range of C2F Network



The black box is the search range relative to inputs' resolution at each resolution level. The matching order of network is in arrow direction.

The Theoretical Receptive Field of GMC

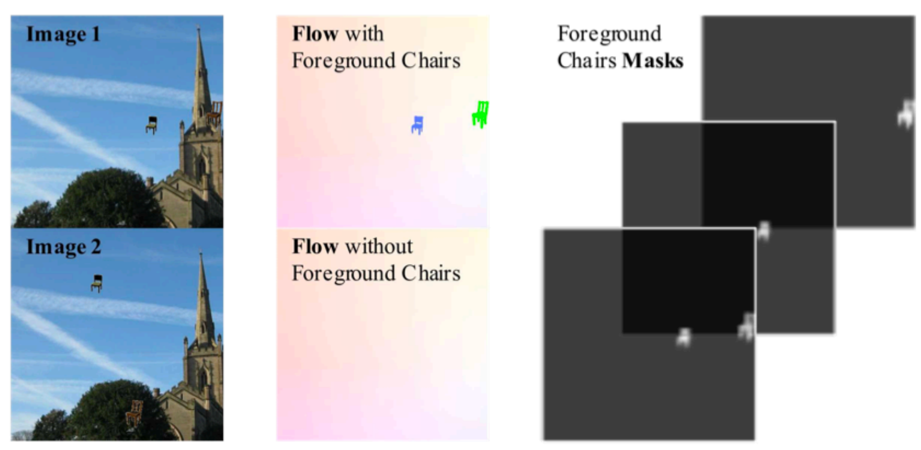


The black box region is the search range relative to inputs' resolution at each resolution level. The convolution order is in arrow direction.

The theoretical receptive field of GMC expands by encoding, and further expands through skip-connection in decoding. As a result, the search range of GMC increases with resolution to whole images and keeps detail information at the same time. This process avoids the dependence of warped features according to up-sampled flows.

SFChairs Dataset

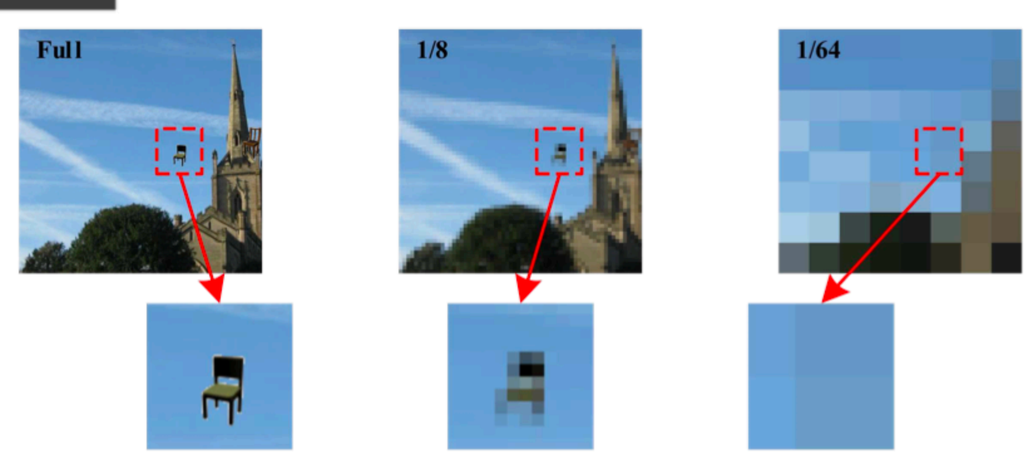
- Dataset split: 10000 examples(with resolution 512x512), 90% training, 10% testing



(a)

(a) is a sample with images, optical flow ground truth, and foreground masks for foreground chairs in SFChairs. All foreground objects are small and fast-moving objects

(b) indicates the foreground chair at different resolution levels, the chair disappears at the lowest resolution level.



(b)

Results

Table 1. AEE ON SFChairs

Models	Training Set			Test Set		
	All	Bg.	Obj.	All	Bg.	Obj.
PWC-Net	(0.62)	(0.27)	(64.54)	0.79	0.27	87.01
HMFlow-G	(0.59)	(0.36)	(45.58)	0.71	0.42	56.03
HMFlow	**(0.39)**	**(0.20)**	**(36.64)**	**0.45**	**0.21**	**44.34**

[a] The **All**, **Bg.** and **Obj.** indicate the AEEs of All image, Background and Foreground Object Regions.
[b] The **HMFlow-G** estimates flows with only GMC's global matching features.

Table 2. AEE ON MPI SINTEL

Methods	Training Set		Test Set		Size
	Clean	Final	Clean	Final	(million)
FlowNetS [11]	(3.66)	(4.44)	6.96	7.76	38.68
FlowNetC [11]	(3.78)	(5.28)	6.85	8.51	39.18
FlowNet2 [12]	(1.45)	(2.01)	4.16	5.74	162.52
SPyNet [13]	(3.17)	(4.32)	6.64	8.36	1.20
LiteFlowNet [14]	**(1.35)**	**(1.78)**	4.54	5.38	5.37
PWC-Net [15]	(1.70)	(2.21)	3.86	5.13	9.37
HMFlow	(1.44)	(2.23)	**3.21**	**5.04**	14.27

[a] The **Size** indicates networks' number of parameters in million.

Figure: Error Maps on MPI Sintel

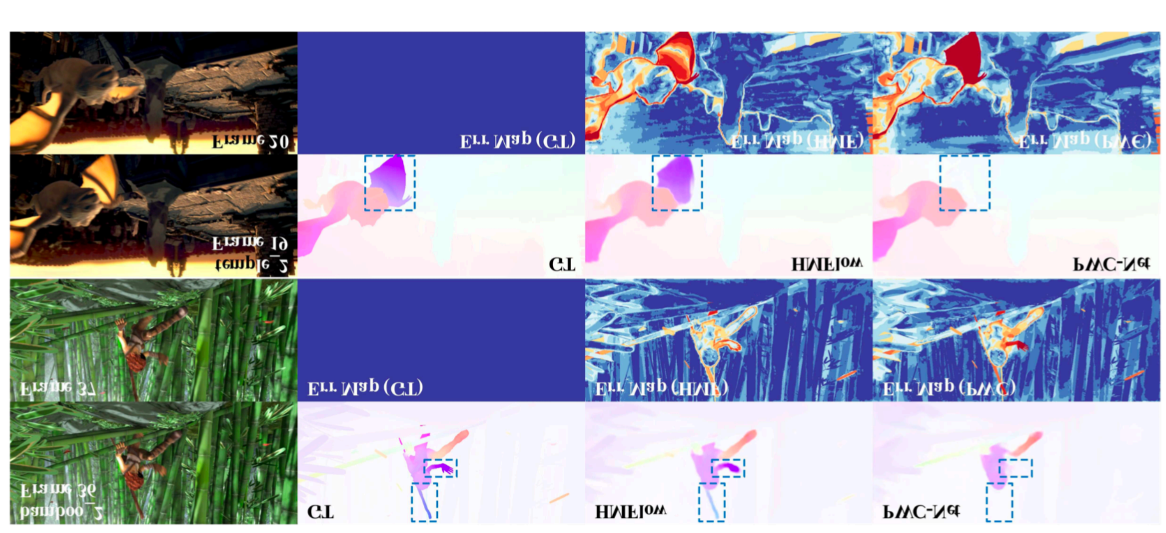
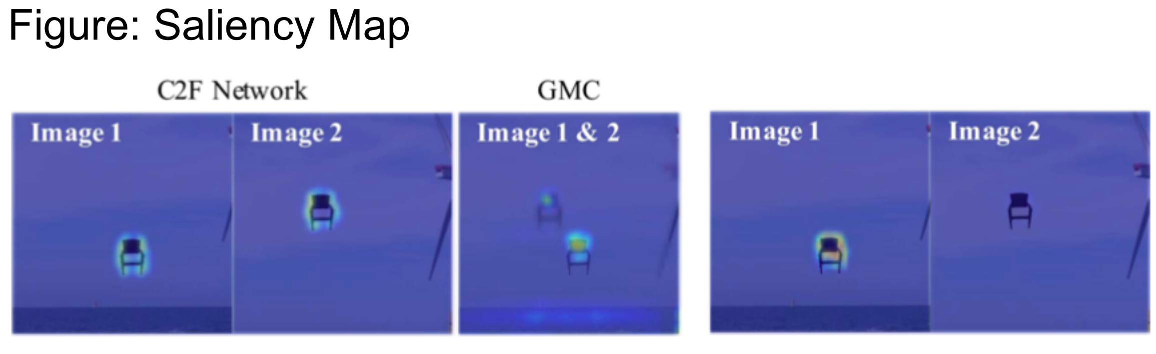
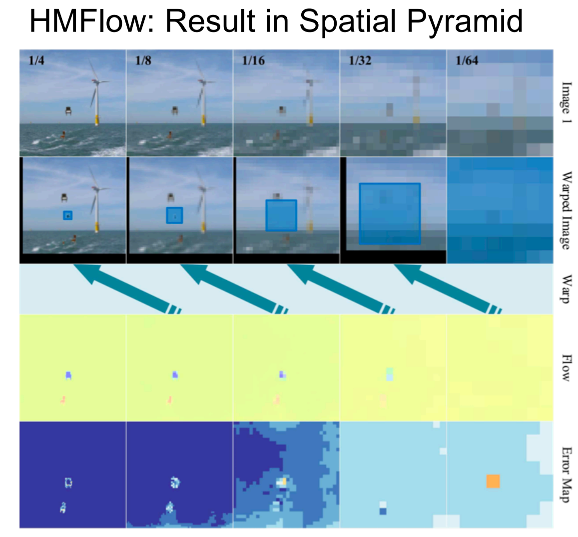


Figure: Saliency Map



HMFlow: Result in Spatial Pyramid



PWCNet: Result in Spatial Pyramid