Optical Flow Guided Feature: A Fast and Robust Motion Representation for Video Action Recognition

Shuyang Sun

The University of Sydney

Trimmed Video Action Recognition

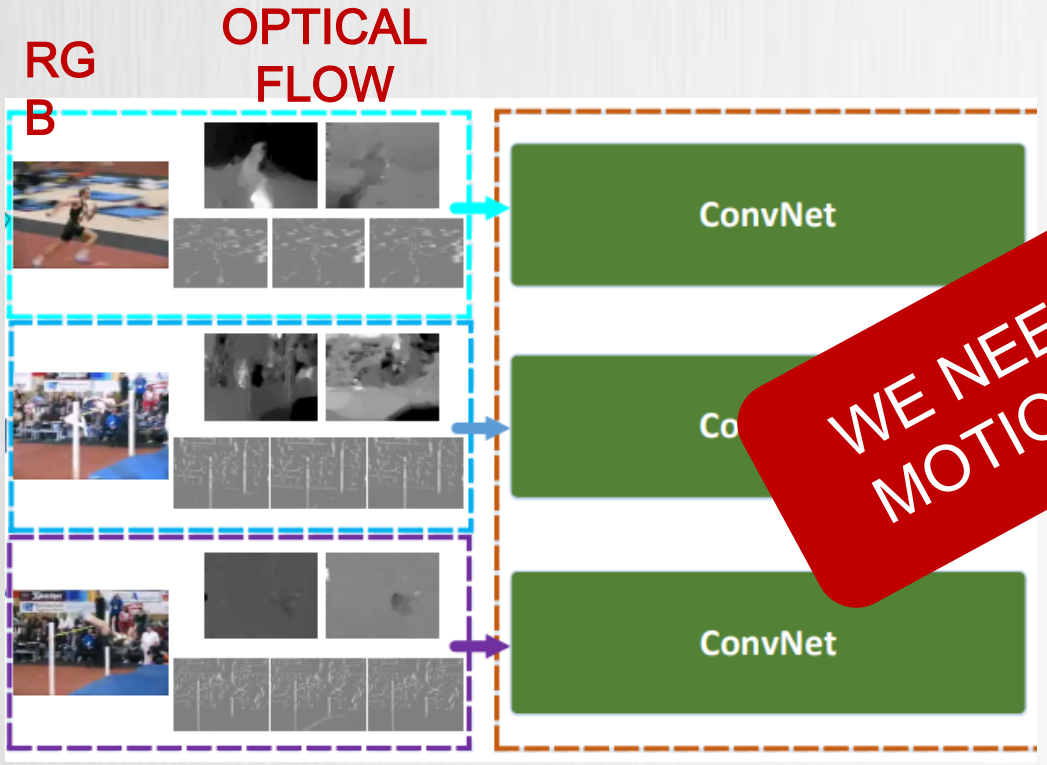
SKATEBOARDIN
G

TRIMMED VIDEO
RECOGNITION

APPLAUDING

IMAGE
RECOGNITION

Problem Settings



TWO-STREAM BASED ARCHITECTURE

Mod	Speed(fps)	Acc.
	680	85.5%
	340	91.0%
Flow	14	87.9%
RGB+Optical Flow	14	94.0%

WE NEED A BETTER MOTION MODALITY!

The extraction time of optical flow comes is extremely time-consuming

Motivation: How to Design?

Hint from TSN: RGB Diff is fast and useful.

Hint from the definition of optical flow:

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t)$$

RGB Diff

$$\frac{\partial I(x,y,t)}{\partial x} v_x + \frac{\partial I(x,y,t)}{\partial y} v_y + \frac{\partial I(x,y,t)}{\partial t} = 0$$

$\{v_x, v_y\}$ = optical flow

Intuitive Inspiration

Coefficient for optical flow:

$$\left\{\frac{\partial I(x,y,t)}{\partial x}, \frac{\partial I(x,y,t)}{\partial y}, \frac{\partial I(x,y,t)}{\partial t}\right\}$$

Experimental Conclusion for coefficients feeding: **Though fast, still not good enough.**

Solution?

Try on feature level!

Optical Flow Guided Feature (OFF): Definition

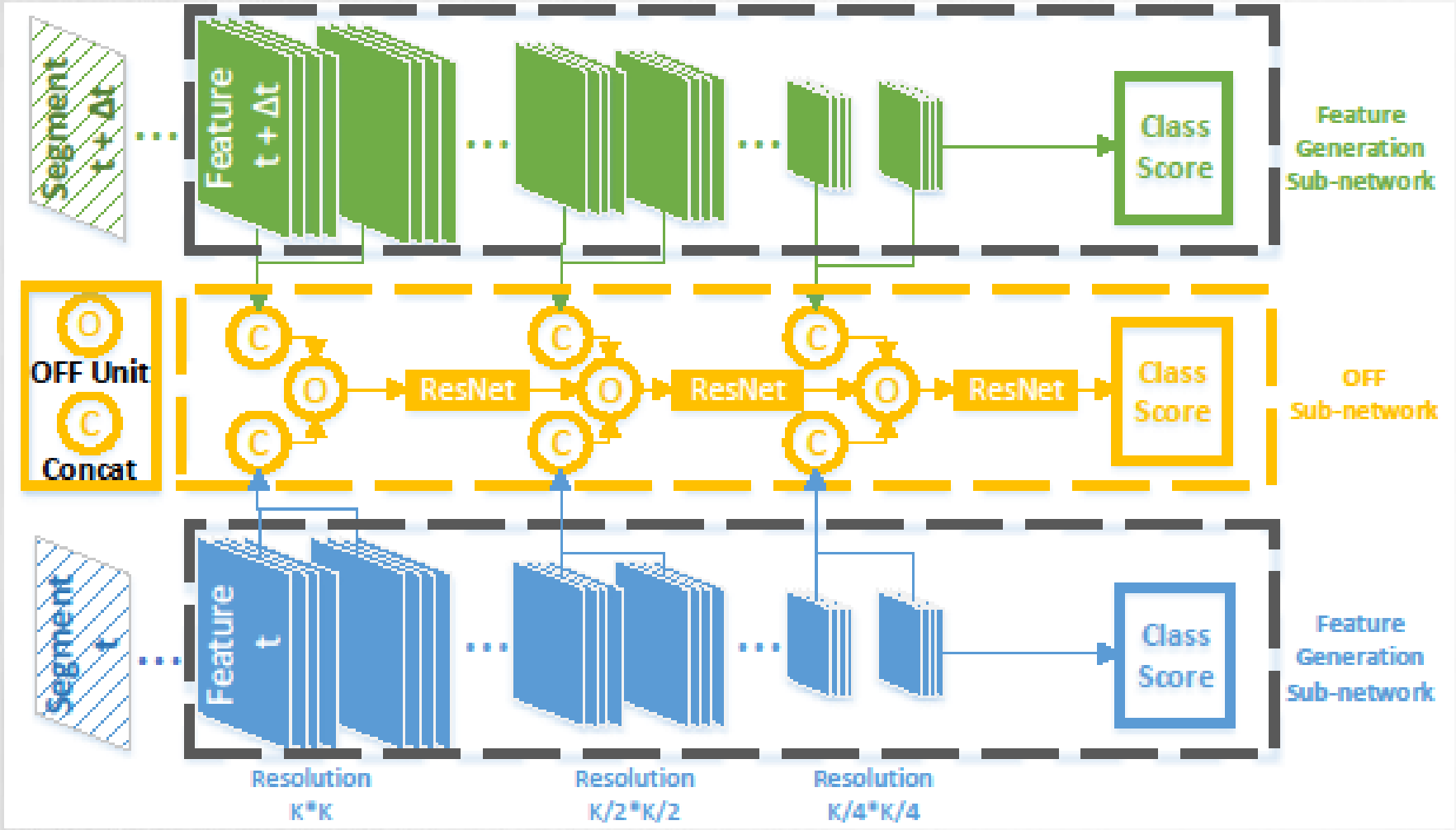
For any differentiable function f:

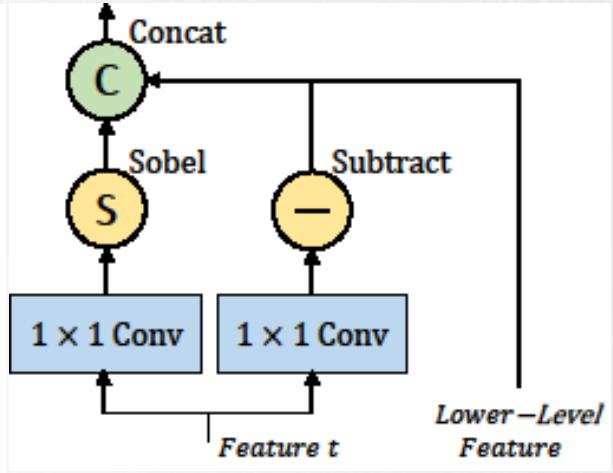
$$f(I(x,y,t);w) = f(I(x+\Delta x, y+\Delta y, t+\Delta t);w)$$

$$\frac{\partial f(I(x,y,t);w)}{\partial x} v_x + \frac{\partial f(I(x,y,t);w)}{\partial y} v_y + \frac{\partial f(I(x,y,t);w)}{\partial t} = 0$$

OFF: $\left\{\frac{\partial f(I(x,y,t);w)}{\partial x}, \frac{\partial f(I(x,y,t);w)}{\partial y}, \frac{\partial f(I(x,y,t);w)}{\partial t}\right\}$

Optical Flow Guided Feature (OFF): Implementation





Optical Flow Guided Feature (OFF): Performance

Method	Speed (fps)	Acc.
TSN(RGB) [43]	680	85.5%
TSN(RGB+RGB Diff) [43]	340	91.0%
TSN(Flow) [43]	14	87.9%
TSN(RGB+Flow) [43]	14	94.0%
RGB+EMV-CNN [50]	390	86.4%
MDI+RGB [3]	<131	76.9%
Two-Stream I3D (RGB+Flow) [5]	<14	93.4%
RGB+OFF(RGB)+ Raw OFF+OFF(Raw OFF)	206	**93.3%**

Table 1. Experimental results of accuracy and efficiency for different real-time video action recognition methods on *UCF-101 over three splits*. Here the notation *Flow* represents the motion modality Optical Flow. Note that our OFF based algorithm could achieve the state-of-the-art performance among real-time algorithms.

RGB	OFF (RGB)	Raw OFF	OFF (Raw OFF)	Flow	OFF (Flow)	Speed (fps)	Acc.
✓						680	85.5%
✓	✓					450	90.0%
✓		✓				340	90.7%
✓	✓	✓				257	92.0%
✓	✓	✓	✓			206	93.0%
✓				✓		14	93.5%
✓	✓			✓		14	95.1%
✓	✓			✓	✓	14	95.5%

Table 2. Experimental results for different modalities using the OFF on *UCF-101 Split1*. Here Flow denotes the optical flow. OFF(*) denotes the use of OFF for the input *. For example, OFF(RGB) denotes the use of OFF for RGB input. The speed here illustrates the time cost for network forward. The results for RGB and RGB + Flow are from [43]. The OFF(RGB) provides a strong 4.5% improvement when fusing with RGB.

Method		
iDT [40]		
Two-Stream [29]		
Two-Stream TSN [4		
Three-Stream TSN [		
Two-Stream+LSTM		
TDD+iDT [41]		
LTC+iDT [37]		
KVMDF [52]		
STP [44]		
STMN+iDT [12]		
ST-VLMPF+iDT [7]		
L^2STM [32]	93.6%	66.2%
Two-Stream I3D [5]	93.4%	66.4%
Two-Stream I3D (with Kinetics 300k) [5]	98.0%	80.7%
Ours	**96.0%**	**74.2%**

Table 4. Performance comparison to the state-of-the-art methods on UCF-101 and HMDB-51 over 3 splits.

1. The performance of the OFF with only RGB inputs is even **comparable** with the two-stream version of other state-of-the-art methods.

2. The OFF is also applicable to other motion representations like optical flow. The product from a motion modality could be regarded as the acceleration representation.

3. The final settings of OFF could achieve state-of-the-art result on UCF-101 and HMDB-51.

Q&A

THANK YOU

更多分享请关注极市平台公众号

更多视觉算法合作请联系小助手