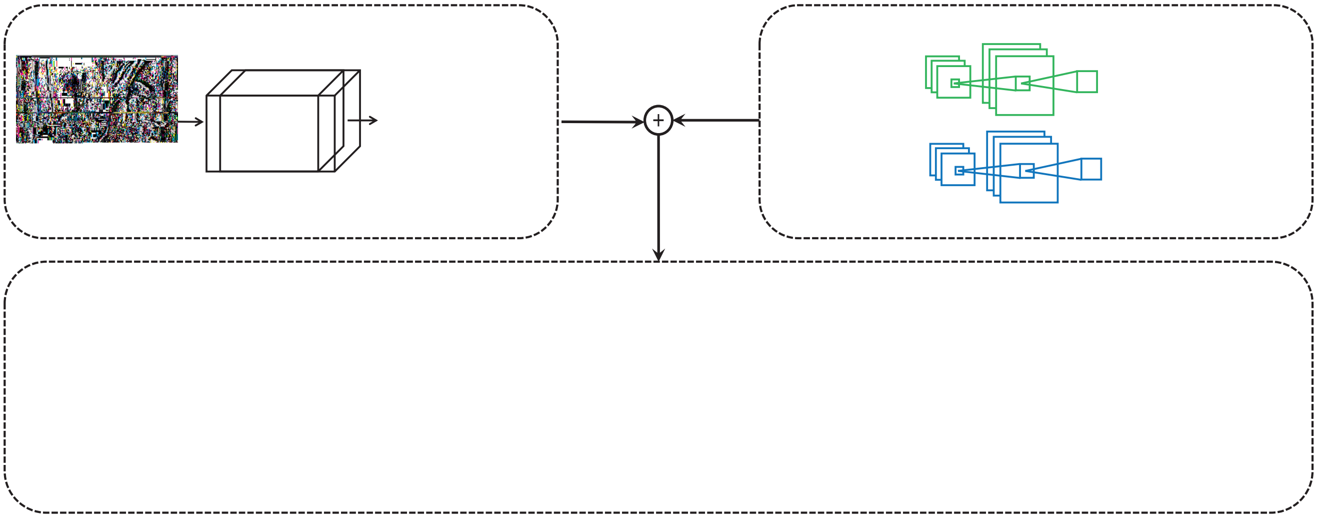
Action Recognition with Trajectory-Pooled Deep-Convolutional Descriptors

07wanglimin@gmail.com, yu.qiao@siat.ac.cn, xtang@ie.cuhk.edu.hk

Abstract

Visual features are of vital importance for human action understanding in videos. This paper presents a new video representation, called (DD)*, which shares the merits of both hand-crafted features [31] and deep-learned features [24]. Specifically, we utilize deep architectures to learn discriminative convolutional feature maps, and conduct trajectory-constrained pooling to aggregate these convolutional features into effective descriptors. To enhance the robustness of TDDs, we design two normalization methods to transform convolutional feature maps, namely spatiotemporal normalization and channel normalization. The advantages of our features come from (i) TDDs are automatically*

11). I . F , , , H , *deep-learned* C B 29, 3D C 9, D C 12, C 24. C 24 - - 31, 32 CF101 H DB51. , . H , fi 14, C .



F 2. **Pipeline of TDD.** DD. : ( ) , ( ) - , ( ) DD. C . G , - .

[14, 25, 28, 41]

[9, 12, 24, 29]. *et al.* [29]

A AC 8
B - fl ,
warped flow. / fl fl ,
H F,
/ DD
, fi - ,
- DD
(4). /
, , *V*,

$$\mathbb{T}(V) = \{T_1, T_2, \cdots, T_K\}, \quad (2)$$

K , T_k
k^{th} :

$$T_k = \{(x_1^k, y_1^k, z_1^k), (x_2^k, y_2^k, z_2^k), \cdots, (x_P^k, y_P^k, z_P^k)\}, \quad (3)$$

(x_p^k, y_p^k, z_p^k) p^{th}
T_k, P $(P = 15)$.
- DD ,

4. Deep Convolutional Descriptors

, ,
trajectory-pooled deep-convolutional descriptor (-
DD), fi - .
- / fi
(C) . ,
C
- . F ,
, DD .

4.1. Convolutional networks

DD C
. , C DD
- C 24
CF101 H DB51.
- C C -
, ,
$(224 \times 224 \times 3)$,
, ,
fl fi $(224 \times 224 \times 2F, F$
fl). ,

	1	1	2	2	3	4	5	5	6	7	8
	7×7	3×3	5×5	3×3	3×3	3×3	3×3	3×3	-	-	-
	2	2	2	2	1	1	1	2	-	-	

(x_p^k, y_p^k, z_p^k) p^{th} -
T_k, r_m m^{th}- 1, $\overline{(\cdot)}$
. $D(T_k, \widetilde{C}_m^a)$ *trajectory-pooled deep convolutional descriptor*,
- C .

Multi-scale TDD extension.
DD
- DD. F ,
fi H G, H F, BH
32×32.
- .
, Ifi -
fi , fl
. -
fl fi .
C - -
F 2. B
- - DD
Ifi , -
fi
r_m E (7) $r_m \times s$, s
, , $s =$
$1/2, 1/\sqrt{2}, 1, \sqrt{2}, 2$.

5. Experiments

, fi
. F ,
DD .

5.1. Datasets

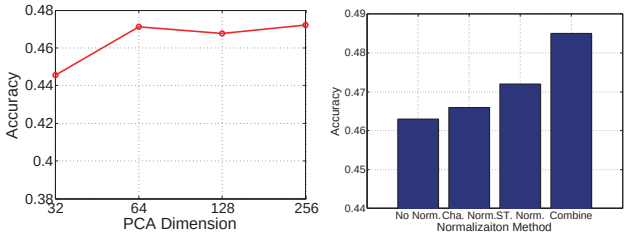
DD ,
, H D-
B51 15 CF101 26 . H DB51
,
6, 766 51 ,
100 .
/ . ,
70 30 . fi
.
CF101 101
100 .

13, 320 ,
25 . /
H 13 11 A
CF101 H DB51, CF101
- C ,
DD H DB51

5.2. Implementation details

Two-stream ConvNets training. C -
I
6 . / CF101 1
- C
C 10
C - (256)
(0.9). F ,
fi 256,
224×224 . I
fl . / - .
4 . F ,
fi CF101 ,
10^{-2}, 10^{-3}
14K , $20K$.
F , 3D
fl fi . / 1 fl
40 C ,
Ifi . F
fl fi
0-255 .
fi , 10
fl fi
Ifi . / CF101 .
A
, ./
0.9 6 0.8 7 .
$224 \times 224 \times 20$ -
fl .
10^{-2} 10^{-3} 50K . I
10^{-4} 70K
90K .

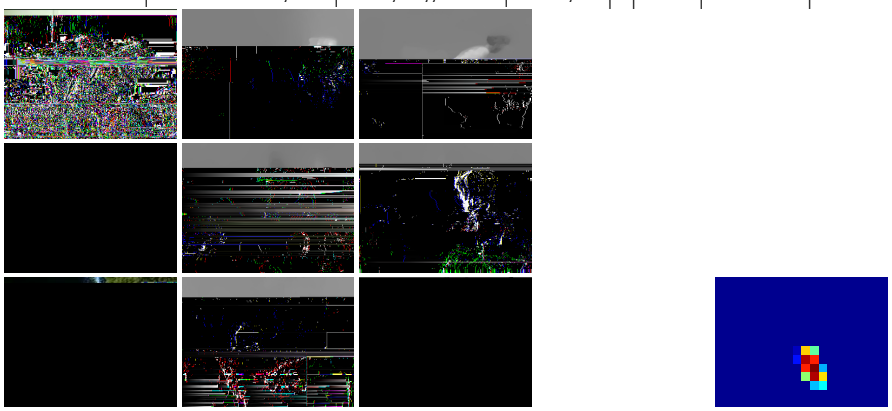
Results of two-stream ConvNets.
, 24, 25
10 fi
. / 71.2%
80.1%
- C 84.7%,
- C 24 (85.6%).


0.48
0.46
0.44
0.42
0.4
0.38
Accuracy
32
64
128
256
PCA Dimension
0.49
0.48
0.47
0.46
0.45
0.44
Accuracy
No Norm.
Cha. Norm.
ST. Norm.
Combine
Normalizaiton Method

F 3.

	C					C				
C	1	2	3	4	5	1	2	3	4	5
	24.1%	33.9%	41.9%	**48.5%**	**47.2%**	39.2%	50.7%	**54.5%**	**51.2%**	46.1%

2. H DB51



( ) GB ( ) F - ( ) F - ( ) - 4 ( ) - 5 ( ) - 3 ( ) - 4

F 4. E , fl fi ,

fl DD

DD

- C . A -

C , DD

DD

DD . / DD

C

4% 2% , .

/ , fl

fi , F 4.

F ,

F ,

DD D , F

3, DD

-

Computational costs. C D , fl

fl . C

C Ifi .

5.5. Comparison to the state of the art

4

H DB51

H DB51		CF101	
I +B / 15	23.0%	I +B / 26	43.9%
35	42.1%	D 12	63.3%
D +B 30	46.6%	D + AD 3	79.9%
D + 3	55.9%	D + 3	83.5%
D +F 31	57.2%	D +F 32	85.9%
D +H 21	61.1%	D +H 21	87.9%
24	59.4%	24	88.0%
DD+F	63.2%	DD+F	90.3%
	65.9%		**91.5%**

4. C DD . /

DD

DD D .

CF101. DD

H DB51 ,

4.8%,

CF101 , 3.5%.

DD

- -

6. Conclusions

(-

DD), -

D

DD .

Acknowledgement

K40
G IDIA C
H K D F -
F C (91320101, 61472410).
B (JC J20120903092050890, J-
C J201206171146144 38, JC J20130402113127496), 100
LCA , G -
(.201001D0104648280).

References

1. J. K. A : A
. *ACM Comput. Surv.*, 43(3):16, 2011. 1

2. H. B , . , . J. . G F:
. *ECCV*, 2006. 3

3. C , . , , . -
. *CVPR*, 2014. 8

4. K. C fi K , A. A.
: D
. *BMVC*, 2014. 6

5. . D B. . H

[37] . . . . . *ECCV*, 2014. 2

[38] . . . . . *ACCV*, 2012. 7

[39] G. . . . . J. . G . A . . . . . *ECCV*, 2008. 2

[40] C. . . . H. B . A . . . . . *29th DAGM Symposium on Pattern Recognition*, 2007. 6

[41] . D. . . F . . . . . *ECCV*, 2014. 3, 4, 5

[42] J. . B. . . . . *ICCV*, 2013. 2