University of Dundee

Sequential Recognition of Manipulation Actions Using Discriminative Superpixel Group Mining

Tianjun Huang, Stephen McKenna
Computer Vision and Image Processing Group
School of Science and Engineering
University of Dundee, UK

sicsa* The Scottish Informatics & Computer Science Alliance

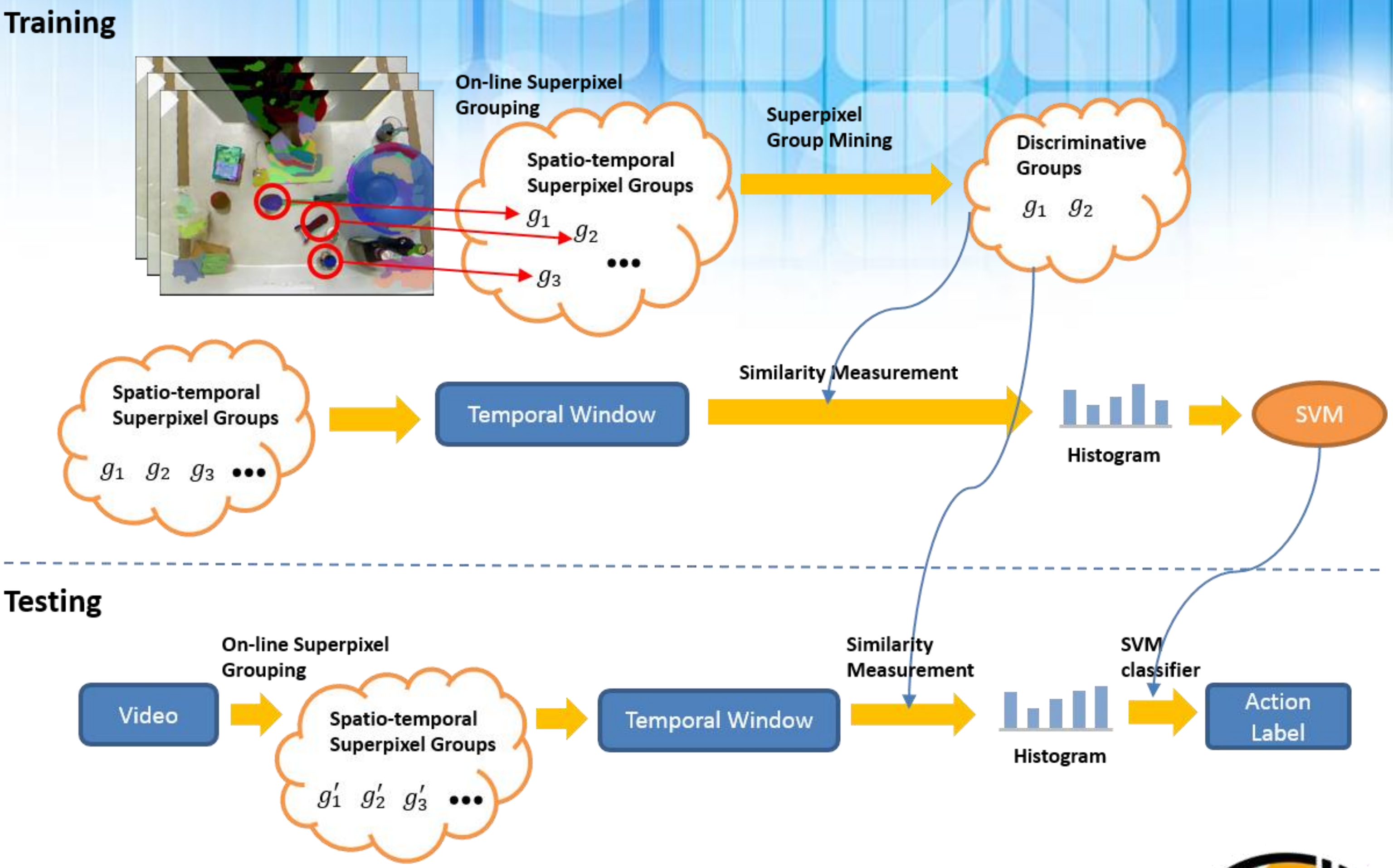
CVIP DUNDEE

50 Salads Dataset

NULL
add_oil
give_pepper
dress_salad
mix_dressing
mix_ingredients
peel_cucumber
cut_into_pieces
place_into_bowl
serve_salad

https://www.youtube.com/watch?v=9r9xyw7fmTg

Overview of The Method



Training

On-line Superpixel Grouping

Spatio-temporal Superpixel Groups

g_1 g_2 g_3 •••

Superpixel Group Mining

Discriminative Groups

g_1 g_2

Spatio-temporal Superpixel Groups

g_1 g_2 g_3 •••

Temporal Window

Similarity Measurement

Histogram

SVM

Testing

Video

On-line Superpixel Grouping

Spatio-temporal Superpixel Groups

g'_1 g'_2 g'_3 •••

Temporal Window

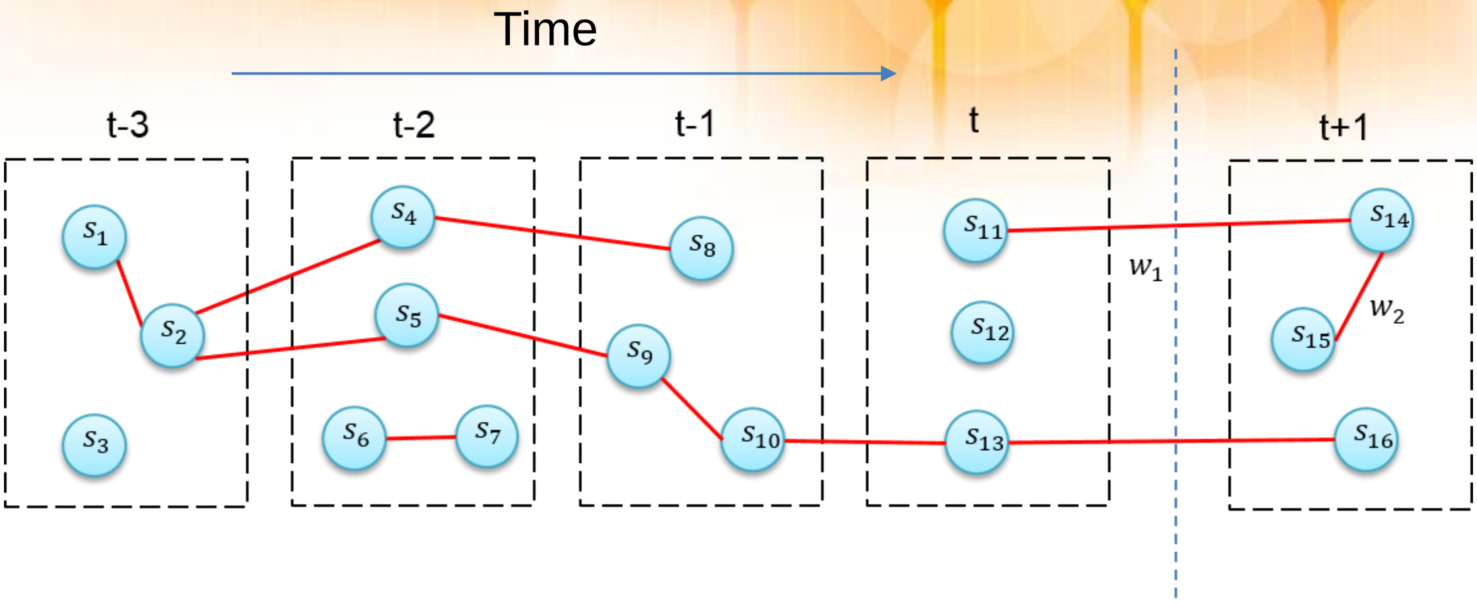
Similarity Measurement

Histogram

SVM classifier

Action Label

Spatio-temporal Superpixel Grouping



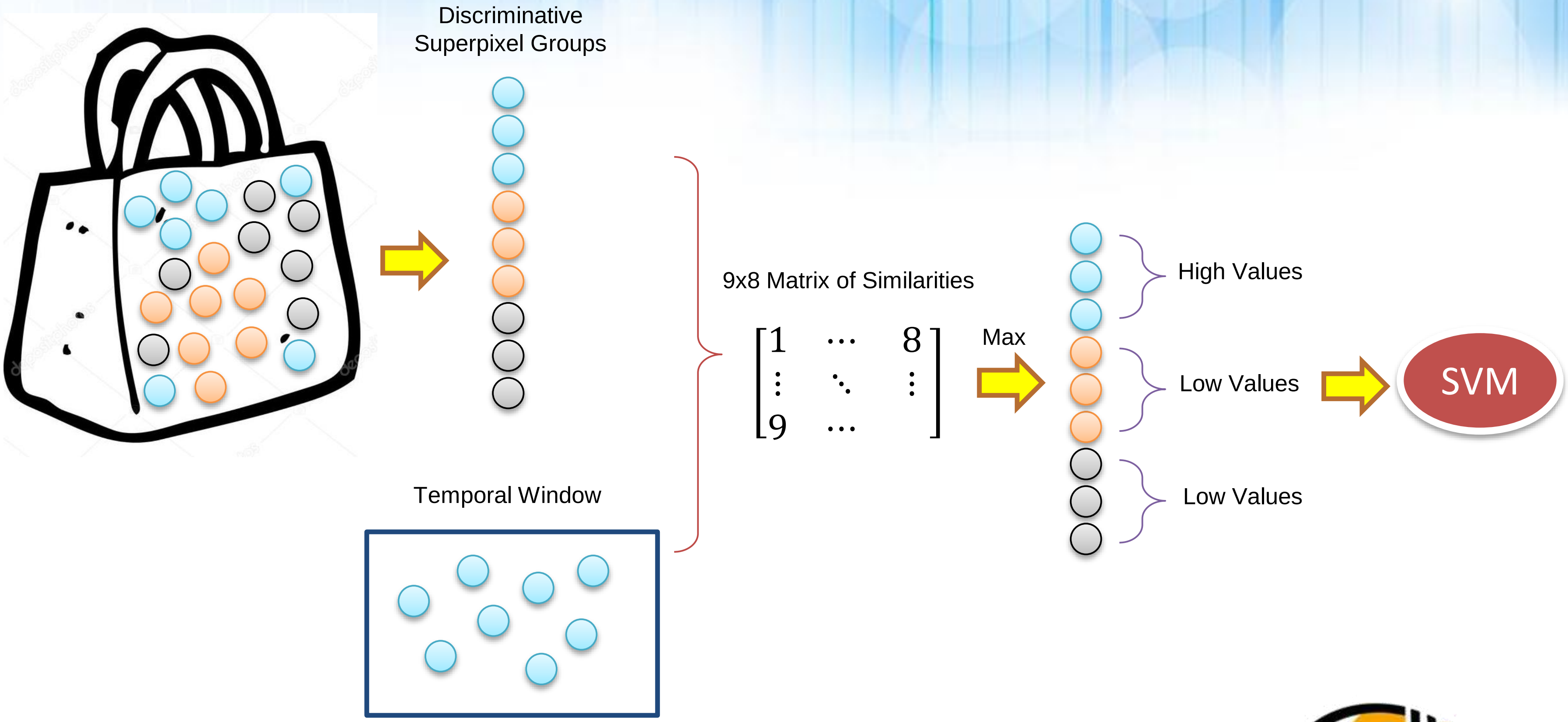
$s_1, s_2, \dots, s_{16}$ denote superpixels in five consecutive frames.

Given a new frame at time t+1, we compute the similarity of nearby superpixels within that frame and the similarity of superpixels in two consecutive frames.

If their similarity is large enough, two superpxiels will be connected and thus belong to the same superpixel group.

University of Dundee

Discriminative Superpixel Group Mining And Temporal Window Representation



Discriminative Superpixel Groups

Temporal Window

9x8 Matrix of Similarities

$$\begin{bmatrix} 1 & \cdots & 8 \\ \vdots & \ddots & \vdots \\ 9 & \cdots & \end{bmatrix}$$

Max

High Values

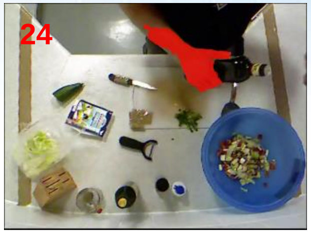
Low Values

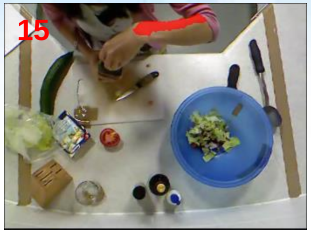
Low Values

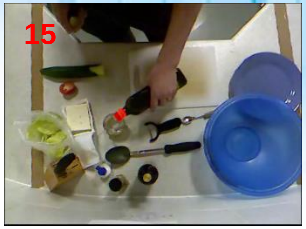
SVM

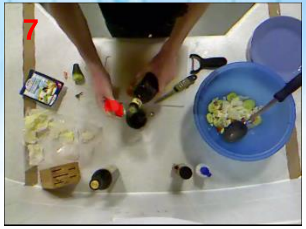
Discriminative Superpixel Groups

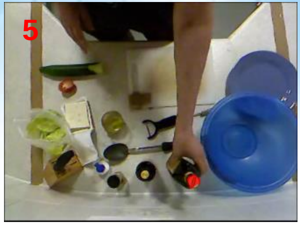
Add Oil



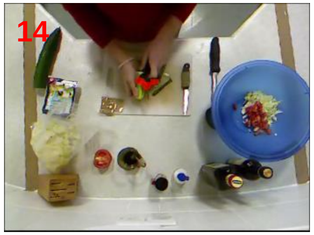


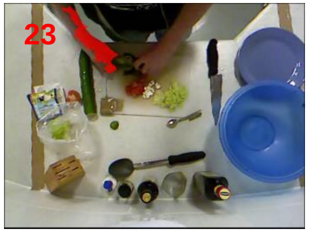


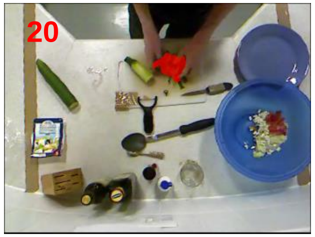


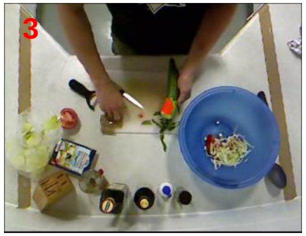


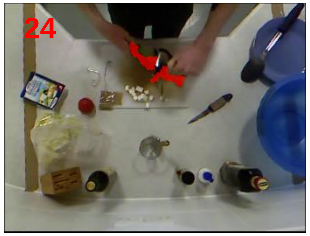
Peer Cucumber



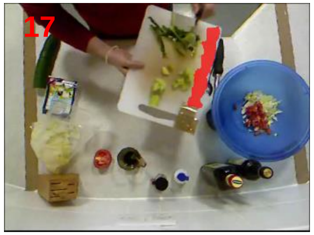


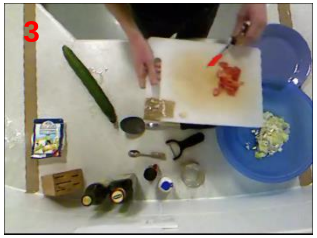


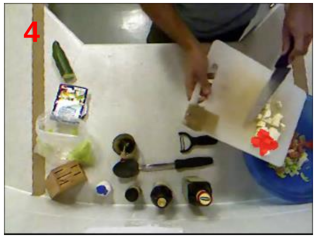


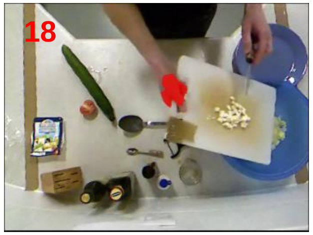


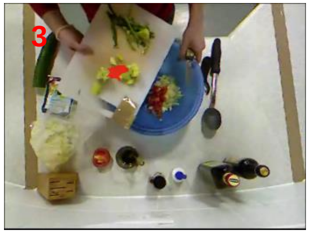
Place Ingredient into Bowl



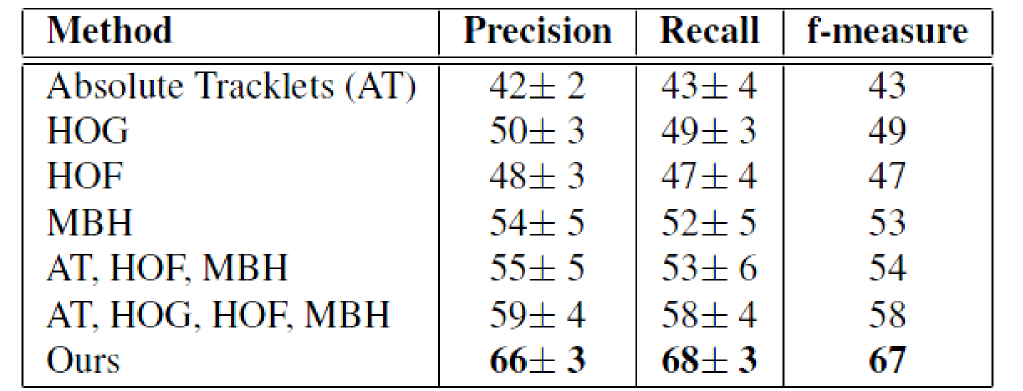






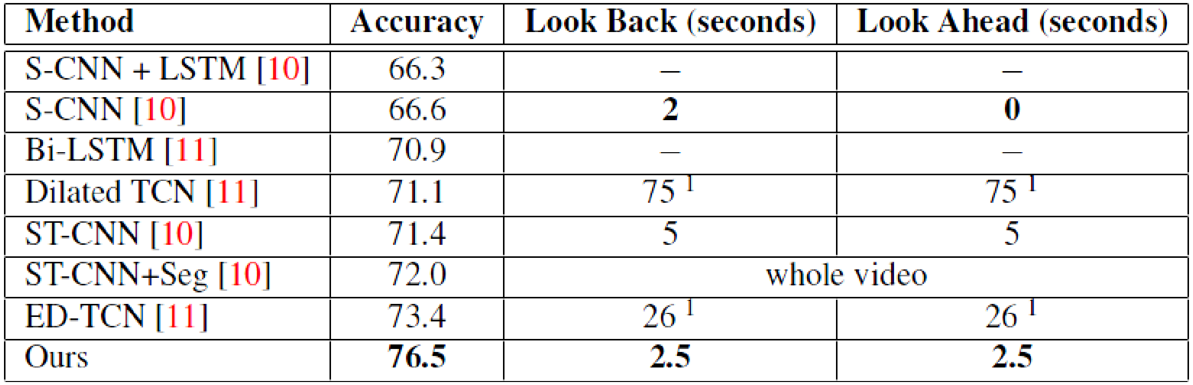


Results



Method	Precision	Recall	f-measure
Absolute Tracklets (AT)	42± 2	43± 4	43
HOG	50± 3	49± 3	49
HOF	48± 3	47± 4	47
MBH	54± 5	52± 5	53
AT, HOF, MBH	55± 5	53± 6	54
AT, HOG, HOF, MBH	59± 4	58± 4	58
Ours	**66± 3**	**68± 3**	**67**

Comparison with multiple visual features and their combinations (Sebastian Stein and Stephen J. McKenna CVIU 2017).



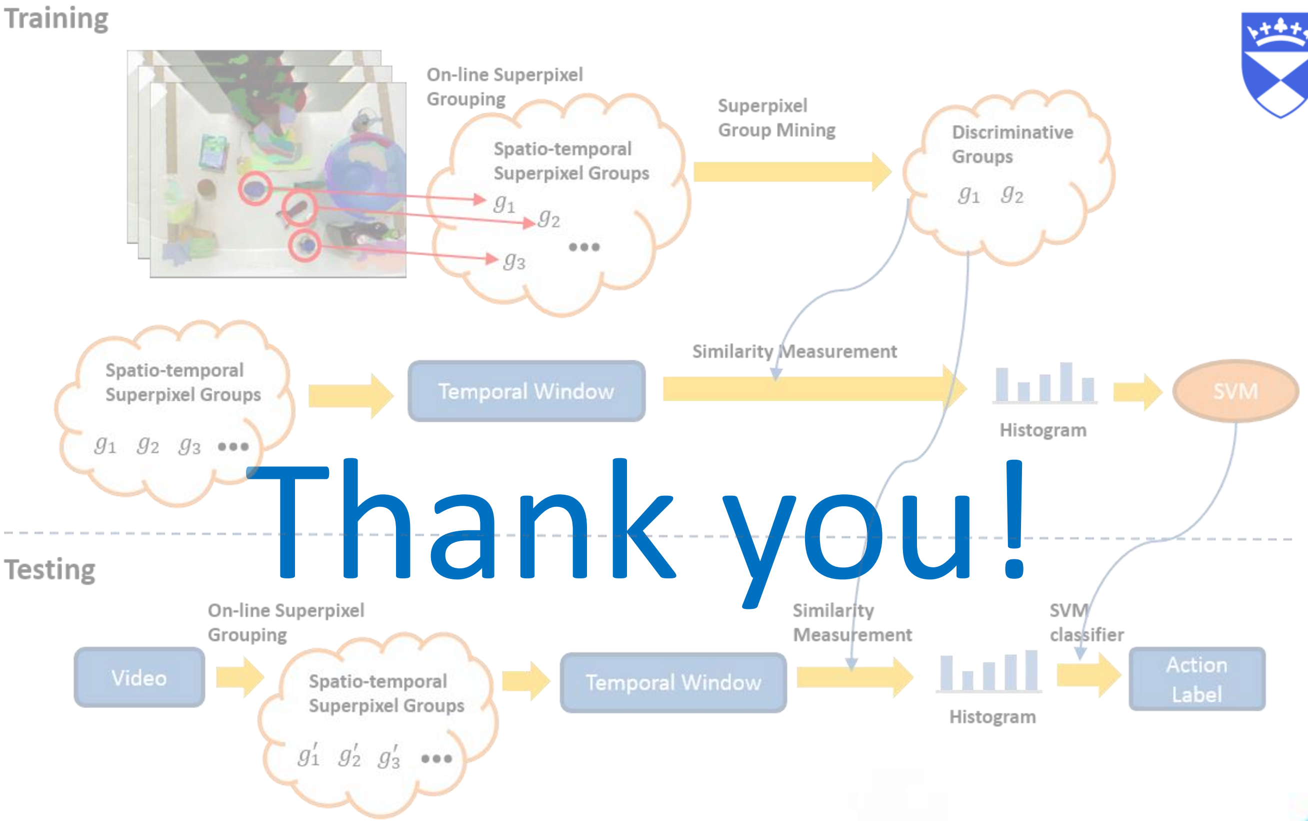
Method	Accuracy	Look Back (seconds)	Look Ahead (seconds)
S-CNN + LSTM [10]	66.3	–	–
S-CNN [10]	66.6	**2**	**0**
Bi-LSTM [11]	70.9	–	–
Dilated TCN [11]	71.1	75 [1]	75 [1]
ST-CNN [10]	71.4	5	5
ST-CNN+Seg [10]	72.0	whole video	
ED-TCN [11]	73.4	26 [1]	26 [1]
Ours	**76.5**	**2.5**	**2.5**

Comparison with deep learning methods (Colin Lea et al. ECCV 2016 [10], CVPR 2017[11]).

Conclusion

- Superpixel groups are able to capture fine-grained motion and object transformation information which is often missing in previous methods.
- Manual annotations for object detections are not required.
- Outperformed methods with comparable temporal windows, and it outperforms CNN methods that use longer temporal windows.
- Method has a directly interpretable representation and can be applied to on-line recognition tasks due to the sequential nature of feature computation.

University of Dundee


Training

On-line Superpixel Grouping

Spatio-temporal Superpixel Groups

g_1 g_2 g_3 •••

Superpixel Group Mining

Discriminative Groups

g_1 g_2

Spatio-temporal Superpixel Groups

g_1 g_2 g_3 •••

Temporal Window

Similarity Measurement

Histogram

SVM

Testing

Video

On-line Superpixel Grouping

Spatio-temporal Superpixel Groups

g'_1 g'_2 g'_3 •••

Temporal Window

Similarity Measurement

Histogram

SVM classifier

Action Label

Thank you!

Tianjun Huang, email: t.huang@dundee.ac.uk
Stephen McKenna, email: stephen@computing.dundee.ac.uk