FEATURE SAMPLING STRATEGIES FOR ACTION RECOGNITION

Youjie Zhou, Hongkai Yu, and Song Wang
University of South Carolina

ICIP 2017
IEEE International Conference on Image Processing
September 17-20, 2017, Beijing, China

Feature Sampling for Action recognition

- Although dense local spatial-temporal features with bag-of-features representation achieve state-of-the-art performance for action recognition, the huge feature number and feature size prevent current methods from scaling up to real size problems.
- In this work, we investigate different types of feature sampling strategies for action recognition, namely dense sampling, uniformly random sampling and selective sampling. We propose two effective selective sampling methods using object proposal techniques.

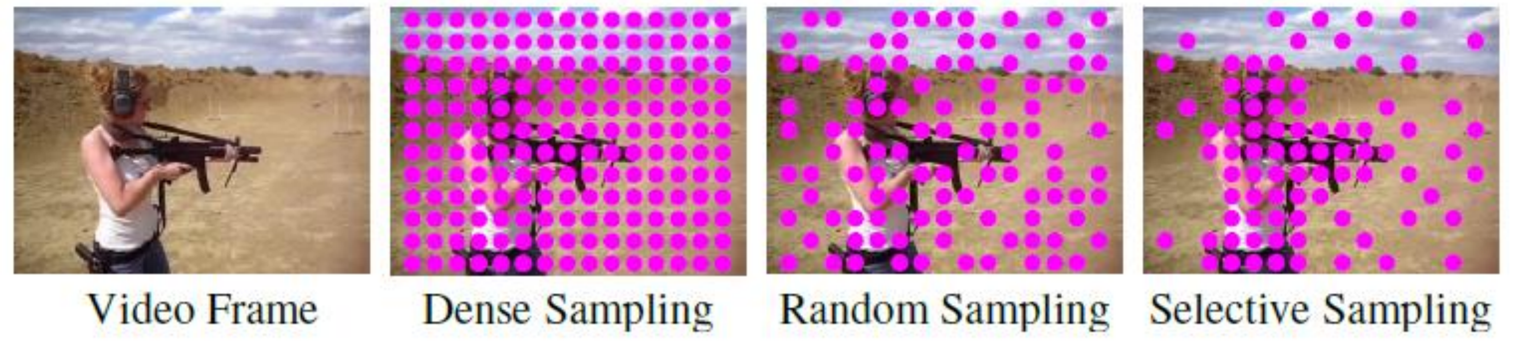

Fig. 1. Different feature sampling methods for action recognition.

Promising Results

Experiments conducted on a large video dataset show that

- we are able to achieve better average recognition accuracy using 25% less features, through one of the proposed selective sampling methods
- even maintain comparable accuracy while discarding 70% features.

Action Recognition Method

- Dense trajectory (DT) features
- Improved dense trajectory (iDT) features
- Linear SVM for classification

Feature Sampling Strategies:

- Random sampling
- Selective sampling via EdgeBox proposals
- Selective sampling via FusionEdgeBox proposals

FusionEdgeBox vs EdgeBox

$$s_{\text{fusion}} = \alpha s_{\text{obj}} + \beta s_{\text{motion}}$$

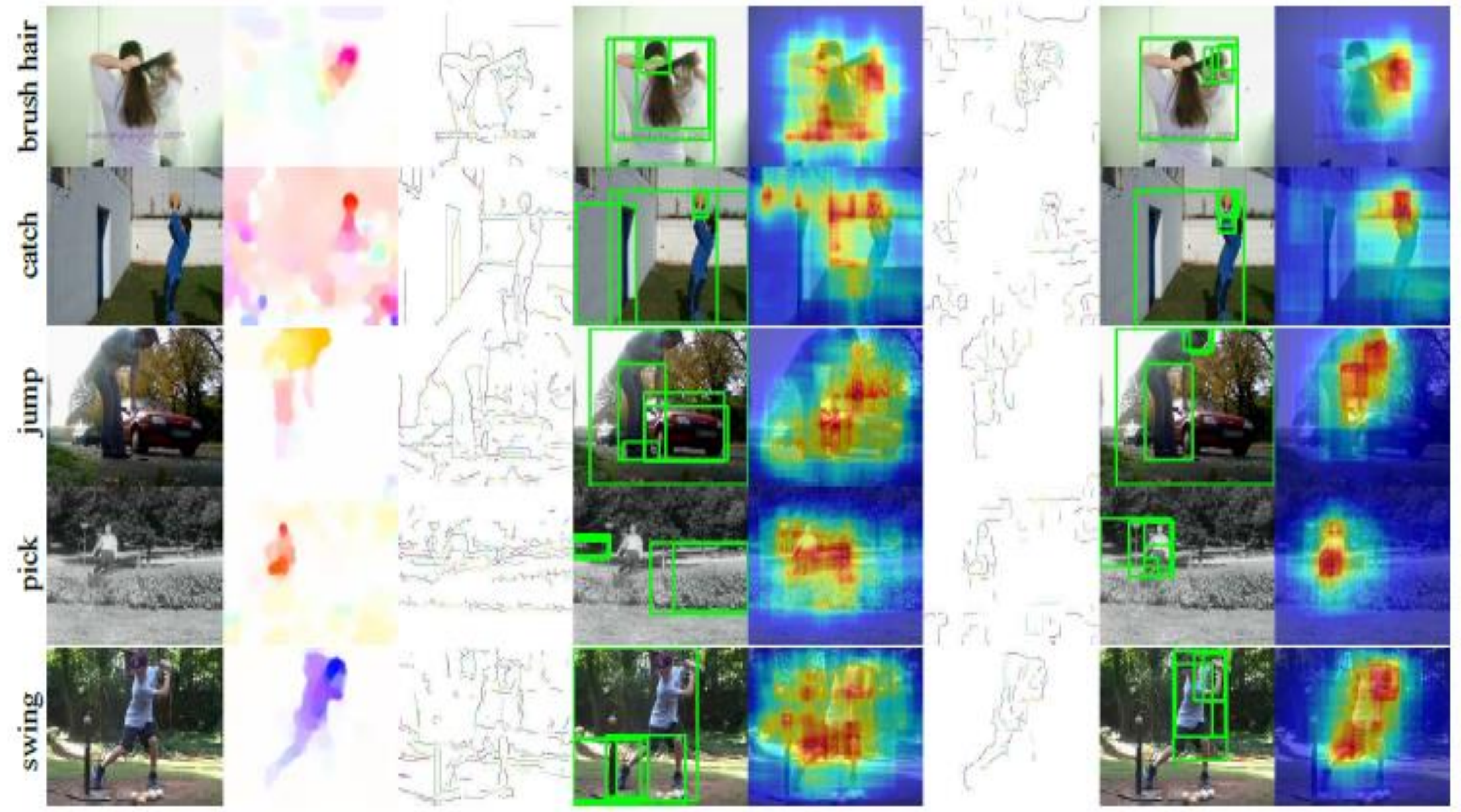

Fig. 2. Illustration of selective sampling methods via object proposal algorithms. From left to right, the original video frame, dense optical flow field, estimated object boundaries, top 5 scoring boxes generated by EdgeBox, saliency map constructed using EdgeBox proposals, estimated motion boundaries, top 5 scoring boxes generated by FusionEdgeBox, saliency map constructed using FusionEdgeBox.

Dataset

We have conducted experiments on one publicly available video datasets, namely J-HMDB [20], which consists of 920 videos of 21 different actions.

Experimental Results

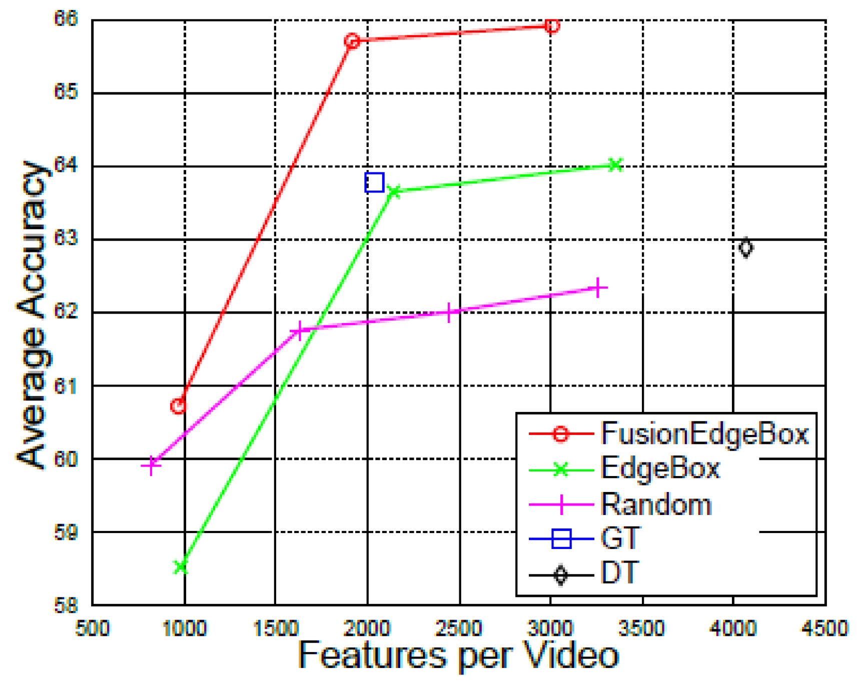

Fig. 3. Average accuracies using the DT feature.

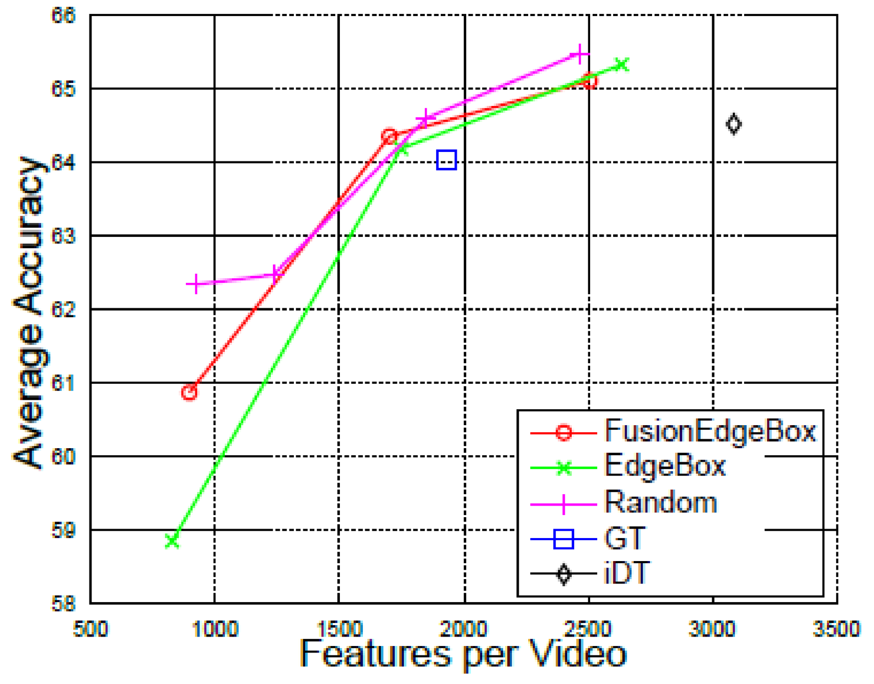

Fig. 4. Average accuracies using the iDT feature.

Method		J-HMDB	Memory (GB)
Dense Trajectory [4]		62.88%	5.4
Improved Dense Trajectory [5]		64.52%	4.2
Peng *et al.* [22] w/ iDT		69.03%*	4.2
Gkioxari *et al.* [23]		62.5%	-
Nie *et al.* [24]		61.2%	-
Discard 20% ∼ 25% features			
DT	Random	62.33%	4.3
	EdgeBox	65.33%	4.5
	FusionEdgeBox	**65.91%**	4.0
iDT	Random	**65.49%**	3.4
	EdgeBox	65.32%	3.6
	FusionEdgeBox	65.11%	3.5
Discard 70% ∼ 80% features			
DT	Random	59.90%	1.1
	EdgeBox	58.51%	1.4
	FusionEdgeBox	**60.71%**	1.4
iDT	Random	**62.34%**	1.3
	EdgeBox	58.85%	1.2
	FusionEdgeBox	60.87%	1.3

Table 1. Comparison to other methods in terms of average accuracy and feature size. * It leverages an advanced feature encoding technique, stacked Fisher vector.

Code is available:
https://github.com/z24/FusionEdgeBox