Good Practices for Learning Video Concept Detectors from Social Media

Semantic Indexing with No Annotations Task

SVETLANA KORDUMOVA, XIRONG LI, CEES G.M. SNOEK

UNIVERSITY OF AMSTERDAM

ACKNOWLEDGEMENT

This research is supported by the STW STORY project, the Dutch national program COMMIT, the National Natural Science Foundation of China (No. 61303184), the Basic Research funds in Renmin University of China from the central government, and by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center contract number D11PC20067. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/ NBC, or the U.S. Government.

LEARN DETECTORS FROM SOCIAL MEDIA

The potential of harvesting training data from the web was recognized by many

Tagged Images

A. Ulges *et al.* ICVS 2008 →
← A. Setz *et al.* ICME 2009
← J. Fan *et al.* CVPR 2010
Z. Wang *et al.* CVPR 2010 →
G. Li *et al.* ICMR 2011 →
← S. Zhu *et al.* TMM 2012
A. Ulges *et al.* ICMR 2012 →
← X. Li *et al.* TMM 2013

Tagged Videos

LEARN DETECTORS FROM SOCIAL MEDIA

Research Question 1:

What visual tagging source is most suited for selecting training examples for learning video concept detectors?

POSITIVE EXAMPLES

Not all images tagged with *Car* contain a car

POSITIVE EXAMPLES

X. Li *et al.* TMM 2009

D. Liu *et al.* WWW 2009

S. Zhu *et al.* TMM 2012


r=0, *s*=30.654


r=1, *s*=21.042


r=2, *s*=13.022


r=3, *s*=8.751

. . .


r=n-2, *s*=0.078


r=n-1, *s*=0.031

POSITIVE EXAMPLES

X. Li *et al.* TMM 2009

D. Liu *et al.* WWW 2009

S. Zhu *et al.* TMM 2012

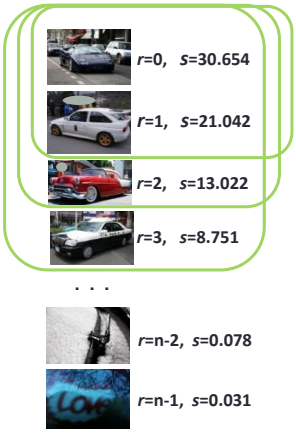
Only ranking

No selection

POSITIVE EXAMPLES

Ulges *et al.* CIVR '08

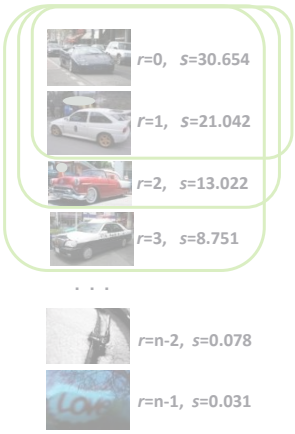
Cross validate different selections



POSITIVE EXAMPLES

Ulges *et al.* CIVR '08

Cross validate different selections



Comes with the expense of manually annotated validation set

POSITIVE EXAMPLES

Calculate cut-off

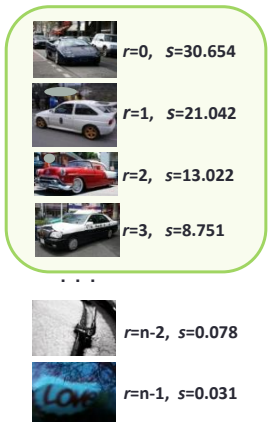
We introduce a binary random variable *y*

y=1 means visual example *x* is positive and *0* otherwise

Bayesian decision:

$$\begin{cases} x \text{ is selected}, if \ \dfrac{p(y=1|x)}{p(y=0|x)} > 1 \\ \text{unselected, otherwise} \end{cases}$$

Kordumova *et al.* CBMI 2013



POSITIVE EXAMPLES

Research Question 2:

What strategy should be used for selecting positive examples from tagged sources when learning video concept detectors?

NEGATIVE EXAMPLES

Common strategy

Random selection of images or videos not tagged with the concept name

A. Ulges *et al.* ICVS 2008

A. Setz *et al.* ICME 2009

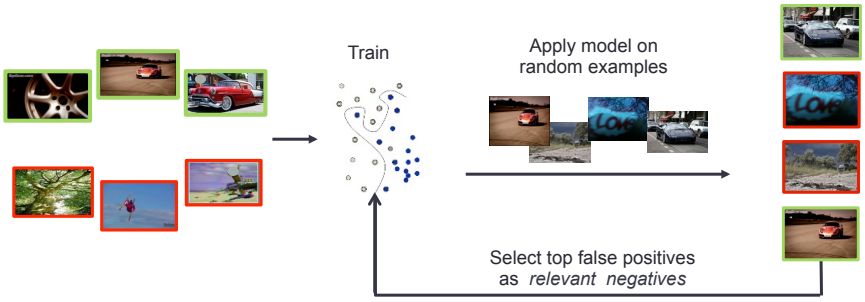
S. Zhu *et al.* TMM 2012

G. Li *et al.* ICMR 2011

A. Ulges *et al.* ICMR 2012

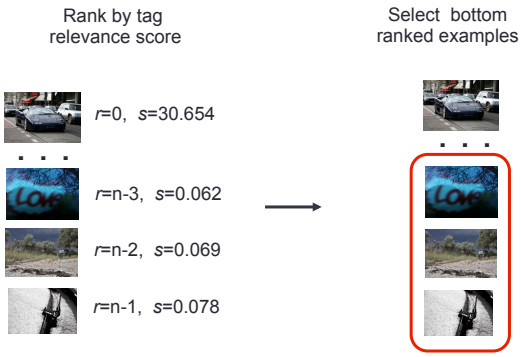
NEGATIVE EXAMPLES

Li *et al.* TMM 2013 – Negative bootstrap



NEGATIVE EXAMPLES

Inspired by R.Yan *et al.* MM 2003 – Pseudo negative



NEGATIVE EXAMPLES

Research Question 3:

What strategy should be used for selecting negative examples from tagged sources when learning video concept detectors?

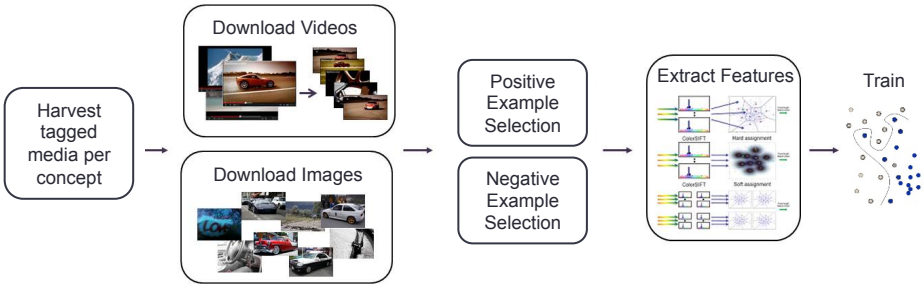
EXPERIMENTS

Experiment 1. What source?
1. Tagged Images
2. Tagged Videos

Experiment 2. What positive examples?
1. Random
2. Relevant ad-hoc
3. Relevant cut-off

Experiment 3. What negative examples?
1. Random
2. Pseudo negative
3. Negative bootstrap

SYSTEM OVERVIEW



RESULTS EXPERIMENT 1

Concept	Tagged Videos	Tagged Images
Animal	0.078	**0.122**
Beach	0.158	**0.359**
Building	0.334	**0.500**
Car	0.157	**0.230**
Child	0.069	**0.118**
City	0.064	**0.131**
Face	0.496	**0.606**
Hand	0.114	**0.175**
Landscape	0.207	**0.567**
Mountain	0.048	**0.516**
Ocean	0.129	**0.481**
Outdoor	**0.816**	0.722
Plant	0.188	**0.270**
Road	0.186	**0.427**
Sky	0.258	**0.621**
Snow	0.063	**0.273**
Sport	0.129	**0.149**
Street	0.099	**0.183**
Tree	0.470	**0.693**
Vehicle	0.221	0.351
mAP	**0.214**	**0.375**

Tagged Images are better source for 19 out of 20 concepts, and 16% better in terms of overall MAP.

RESULTS EXPERIMENT 1

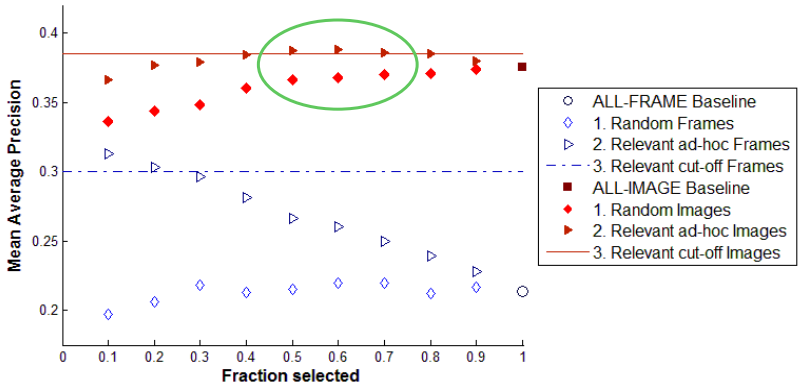
Concept	Tagged Videos	Tagged Images
Animal	0.078	**0.122**
Beach	0.158	**0.359**
Building	0.334	**0.500**
Car	0.157	**0.230**
Child	0.069	**0.118**
City	0.064	**0.131**
Face	0.496	**0.606**
Hand	0.114	**0.175**
Landscape	0.207	**0.567**
Mountain	0.048	**0.516**
Ocean	0.129	**0.481**
Outdoor	**0.816**	0.722
Plant	0.188	**0.270**
Road	0.186	**0.427**
Sky	0.258	**0.621**
Snow	0.063	**0.273**
Sport	0.129	**0.149**
Street	0.099	**0.183**
Tree	0.470	**0.693**
Vehicle	0.221	0.351
mAP	**0.214**	**0.375**

Tagged Images are better source for 19 out of 20 concepts, and 16% better in terms of overall MAP.

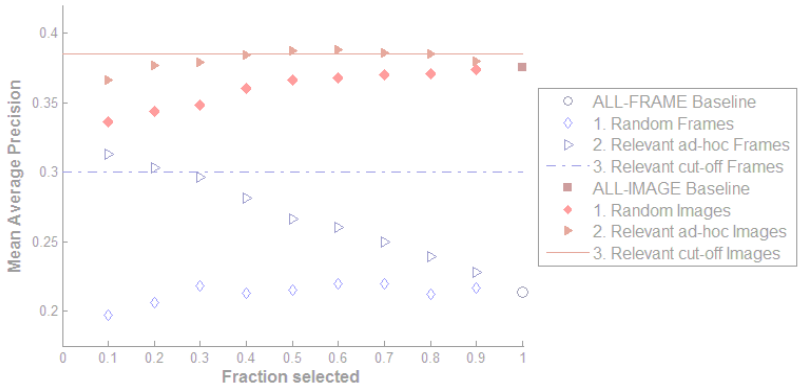
Good practice 1.

Tagged Images are a better source compared to Tagged Videos for learning video concept detectors

RESULTS EXPERIMENT 2



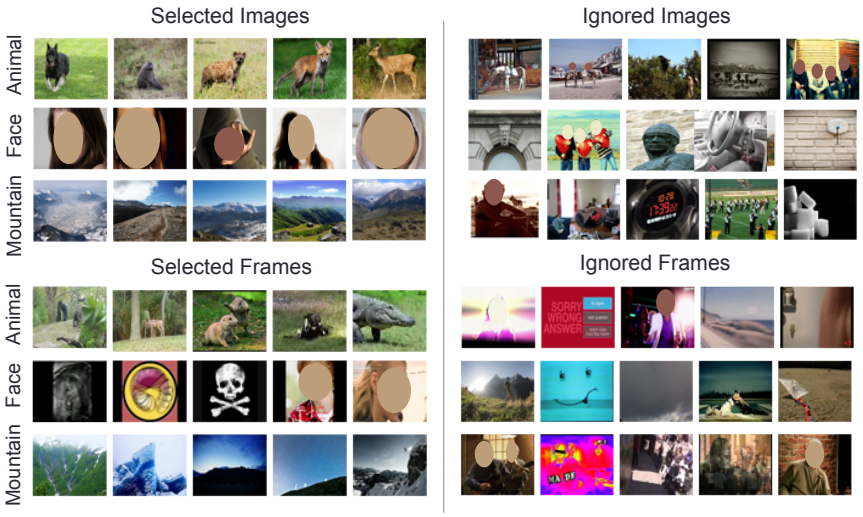
RESULTS EXPERIMENT 2



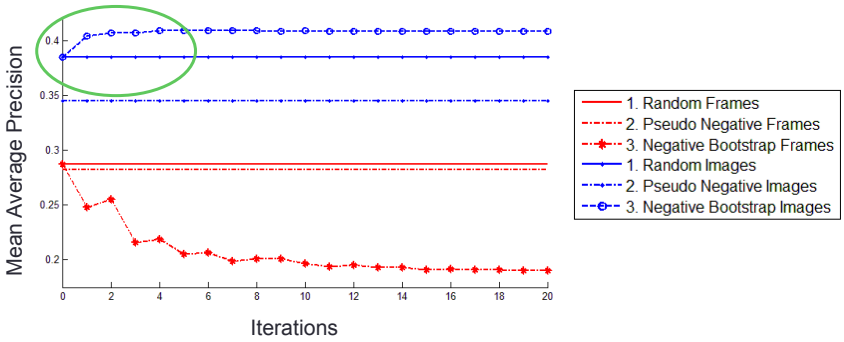
Good practice 2:

For learning video concept detectors from social media,
as positive examples use relevant cut-off selection of tagged images.

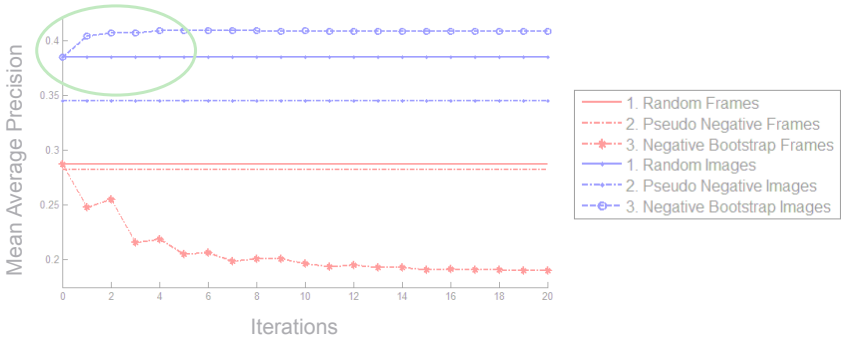
RESULTS EXPERIMENT 2



RESULTS EXPERIMENT 3



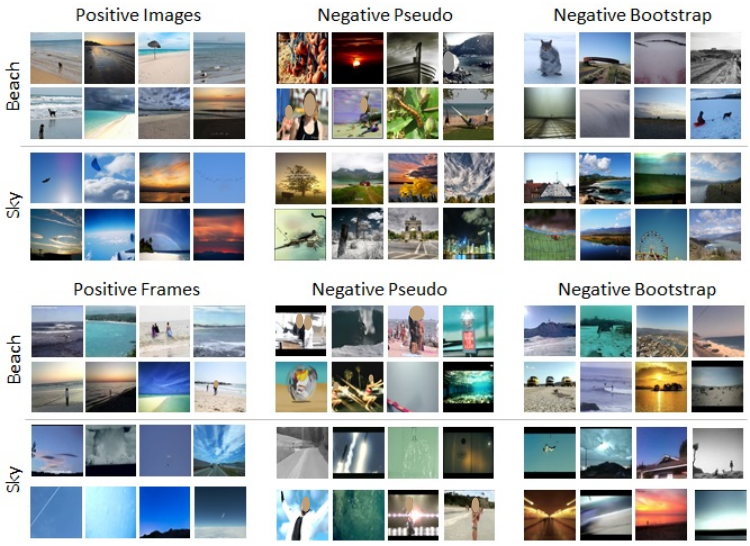
RESULTS EXPERIMENT 3



Good practice 3:

For learning video concept detectors from social media,
use bootstrapping of relevant negatives from tagged images

RESULTS EXPERIMENT 3



TRECVID 2013 SIN NO ANNOTATION

Three good practices

1. Tagged images as a source
2. Relevant cut-off for positive examples from tagged images
3. Negative bootstrap of tagged images

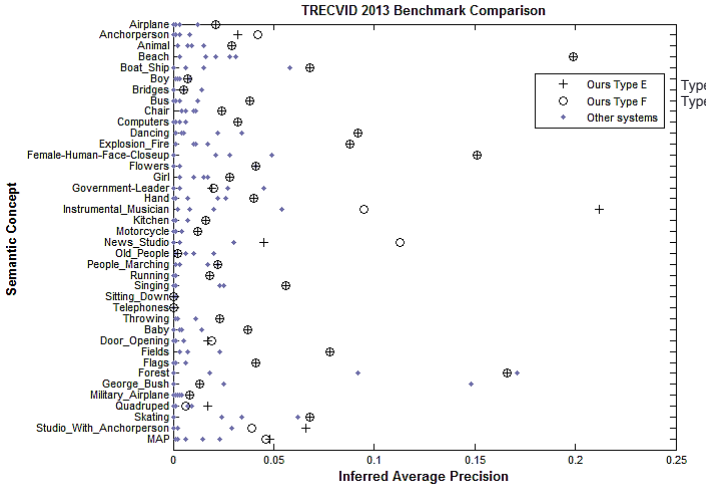
Implementation details

Multi-frame
Densely sampling with SIFT, RGB-SIFT and T-SIFT descriptors
Fisher vector coding with codebook size 1024
Spatial pyramid 1x1+1x3
Linear Kernel SVM

Types

Type E: Automatically generated queries using Wikipedia anchor text and titles of redirect pages
Type F: Manually created queries

TRECVID 2013 SIN NO ANNOTATION RESULTS



Type E: Automatically generated queries
Type F: Manually created queries

Best score for 29 out of 38 concepts, and best overall mAP

GOOD PRACTICES FOR LEARNING VIDEO CONCEPT DETECTORS FROM SOCIAL MEDIA

Good practice 1.

Tagged images are a better source then tagged videos for learning video concept detectors.

Good practice 2.

Positive examples with relevant cut-off of tagged images show best performance.

Good practice 3.

Relevant negatives are best selected with negative bootstrap of tagged images.