

# SVD: A Large-Scale Short Video Dataset for Near-Duplicate Video Retrieval

Qing-Yuan Jiang<sup>†</sup>, Yi He<sup>‡</sup>, Gen Li<sup>‡</sup>, Jian Lin<sup>†</sup>, Lei Li<sup>‡</sup> & Wu-Jun Li<sup>†</sup>

ByteDance

<sup>†</sup>Department of Computer Science and Technology, Nanjing University, Nanjing, China. & <sup>‡</sup>ByteDance AI Lab, Beijing, China.

## Introduction

### Near-Duplicate Video Retrieval (NDVR)

•NDVR: retrieve visually similar near duplicate videos (NDV) from database.

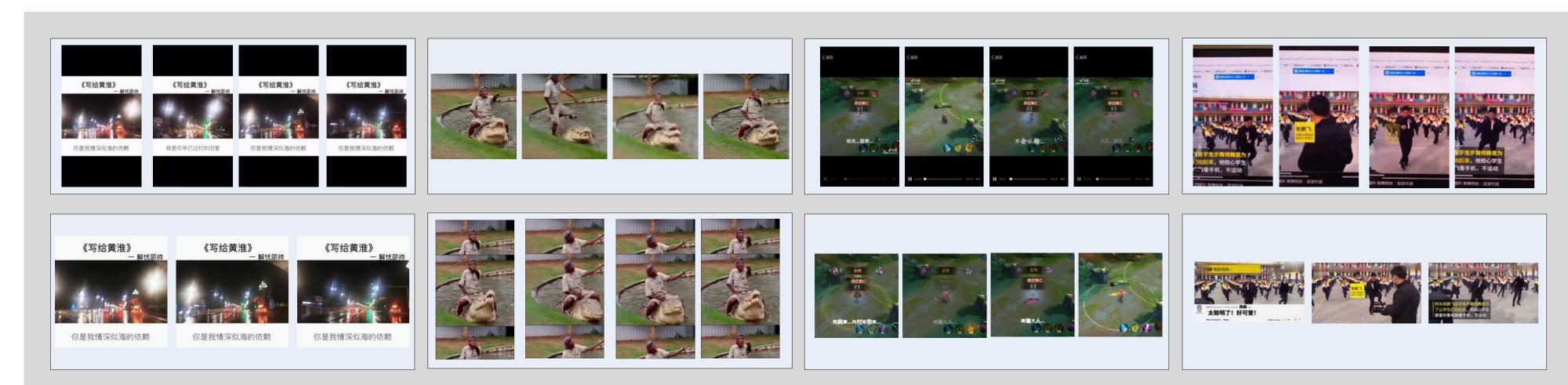


# Existing NDVR datasets

- •CCWEB [MM'2007], UQ\_VIDEO [MM'2011], VCDB [ECCV'2014], MUS-CLE\_VCD [2007], TRECVID [TRECVIDW'2010]
- Existing NDVR datasets:
- -All existing NDVR datasets are for long videos (duration > 60s).
- -Most of them are small-scale and lack of diversity.

#### Short NDVs

- New difficulties for detecting short NDVs:
- -New types of near-duplicates for short videos, e.g., horizontal/vertical screen videos and camera shaking videos.



Some examples in SVD dataset.

-#short NDVs is larger than #long NDVs.

## Short Video Dataset

### Mining Strategies

The SVD dataset is divided into three subsets, i.e., the query set, the labeled set and the probable negative unlabeled set.

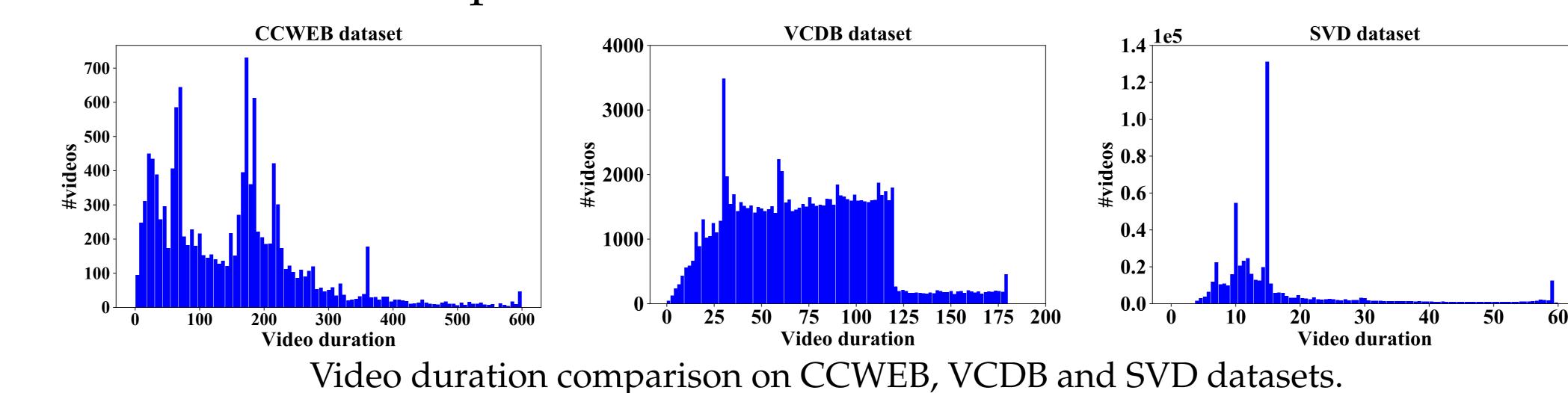
- Query Set: We crawl 1,206 videos to construct query set.
- Labeled Set: Iterative Retrieval, Transformed Retrieval, Feature based Mining.
- Probable Negative Unlabeled Set: Feature based Filtering.

### Short Video Dataset (SVD)

Comparison with existing NDVR datasets.

Comparison to existing NDVR datasets.							
Item	CCWEB	UQ_VIDEO	VCDB	MUSCLE_VCD	TRECVID	SVD	
#query videos	24	24	528	18	11,256	1,206	
#labeled videos	12,790	12,790	528	101	11,503	34,020	
#positive pairs	3,481	3,481	6,139	N/A	N/A	10,211	
#negative pairs	9,311	9,311	0	N/A	N/A	26,927	
#background distraction videos	0	119,833	100,000	0	0	0	
#probable negative unlabeled videos	0	0	0	0	0	526,787	
#total videos	12,814	132,647	100,528	119	22,759	562,013	
Average duration (in second)	151.02	N/A	72.77	3,564.36	131.44	17.33	
Total duration (in hour)	539.95	N/A	2027.60	100	420	2704.96	
Video publically available	$\sqrt{}$	×	$\sqrt{}$	$\sqrt{}$	×	$\sqrt{}$	

Video duration comparison.



#### Transformations

We define four transformations to construct more difficult variants of SVD.

- Temporal Transformation: video speeding.
- **Spatial Transformation:** video cropping, black border insertion, and video rotation.

# Experiment

#### Datasets

- CCWEB, VCDB and SVD.
- ullet Variants of SVD: SVD<sub>Cropping</sub>, SVD<sub>BlackBorder</sub>, SVD<sub>Rotation</sub> and SVD<sub>Speeding</sub>.

#### Benchmarks

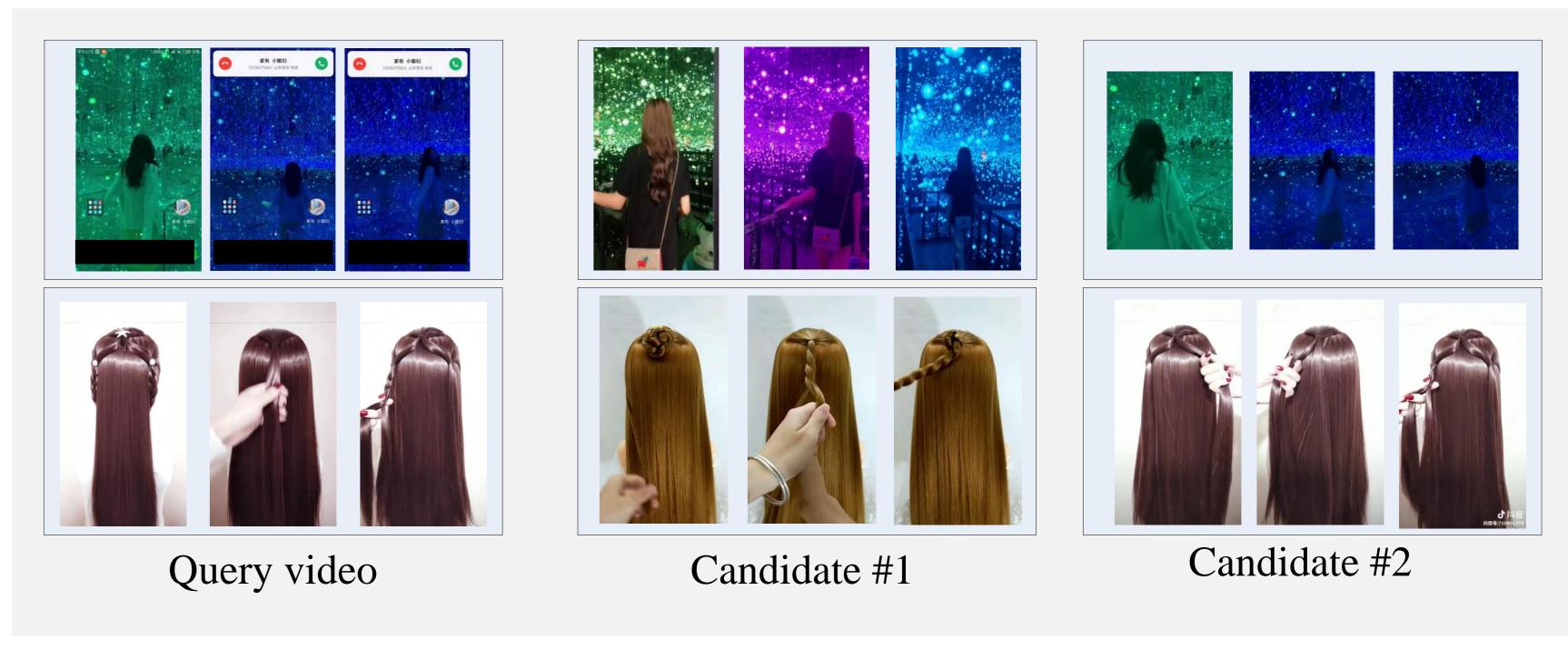
- Real-valued NDVR methods: DML, CNNV, CNNL, and CTE.
- Hashing based NDVR methods: LSH, ITQ, IsoH, and HDML.

#### Accuracy

MAP for real-valued NDVR methods.						
Dataset	DML	CNNL	CNNV	CTE	$\overline{D}$	
CCWEB	97.01	95.47	95.60	90.09	$\overline{C}$	
VCDB	78.98	49.87	45.19	41.42	V	
SVD	78.47	55.55	19.09	50.97	SV	
$\overline{ ext{SVD}_{Cropping}}$	54.06	15.61	6.31	16.48	SV	
$\mathrm{SVD}_{BlackBorder}$	68.17	18.63	6.94	32.66	SV	
${\sf SVD}_{Rotation}$	15.59	0.15	0.22	2.84	SV	
${\sf SVD}_{Speeding}$	76.70	51.80	15.45	16.23	SV	

MAP for hashing based NDVR methods with 32 bits							
Dataset	LSH	ITQ	IsoH	HDML			
CCWEB	83.15	87.14	86.75	90.23			
VCDB	30.88	33.31	33.30	68.92			
SVD	28.36	30.14	30.88	31.59			
$\overline{ ext{SVD}_{Cropping}}$	2.65	4.41	4.01	5.39			
$\mathrm{SVD}_{BlackBorder}$	4.61	7.08	5.58	10.54			
$\mathrm{SVD}_{Rotation}$	0.09	0.43	0.24	1.95			
$\mathrm{SVD}_{Speeding}$	23.56	25.82	26.63	28.60			

# Bad Cases Analysis



Bad cases on SVD dataset.

### Conclusion

- •We introduce a large-scale short video dataset, which contains over 500,000 short videos for NDVR.
- The release of SVD will foster the research of NDVR.

https://svdbase.github.io

