

Motivation:

> Dynamic actions are more complex than visual objects, making traditional attribute/word vector-based zero-shot learning (ZSL) approaches lack of discrimination and scalability:



- Attributes manually specified and highly subjective
- Word Vectors high-dimensional and visual data-irrelevant
- > The superior properties of error-correcting output codes (ECOC) make it favorable for tackling ZSL problems: *e.g., error-correcting* ability, high efficiency, good diversity, and accurate binary classification

Contributions:

- (1) Propose the Zero-Shot ECOC (dubbed ZSECOC): ECOC with the additional capability of zero-shot recognition
- (2) ZSECOC inherits the intrinsic advantages of ECOC as well as overcomes the domain shift problem implicitly and explicitly
- (3) ZSECOC preserves *instance-level* visual data structures in addition to *category-level* semantics
- (4) ZSECOC is systematically evaluated on three realistic video action datasets, i.e., Olympic Sports, HMDB51 and UCF101

Zero-Shot Action Recognition with Error-Correcting Output Codes Jie Qin^{1,2}, Li Liu^{3,4}, Ling Shao⁴, Fumin Shen⁵, Bingbing Ni⁶, Jiaxin Chen¹ and Yunhong Wang¹ ¹Beihang University ²ETH Zurich ³Malong Technologies Co., Ltd. ⁴University of East Anglia ⁶Shanghai Jiao Tong University ⁵University of Electronic Science and Technology of China





UC Code Matri



(2) Preserving category-level semantics (semantics-preserving): $\mathcal{O}_{\rm sp} := \sum s_{ij} ||\mathbf{b}_i - \mathbf{b}_j||_2^2 - \lambda \sum ||\mathbf{b}_i||_2^2 -$

• Optimization: *alternatively* find the solution to each variable

> ZSECOC for unseen categories: • Semantic transfer w/o unseen visual data:

$$\mathbf{B}^{u} = \operatorname{sign}(\mathbf{B}\mathbf{S}^{u}) \quad s_{ij}^{u} = \frac{\langle \phi(y_{i}), \phi(y_{i}) | | \cdot | | \phi(y_{i}) | \cdot | | \cdot | \phi(y_{i}) | | \cdot | | \phi(y_$$

> Zero-shot recognition as Hamming decoding:

 $j^* = \operatorname{argmin} d_{\mathrm{H}}(F(\mathbf{x}^u), \mathbf{b}_i^u)$

$$-1,1\}^{m \times C}$$

$$|\mathbf{P} - \mathbf{R} \mathbf{V}||_{\mathrm{F}}^2$$

$$-\mathbf{b}_{j}||_{2}^{2}$$



Experiments:

- Dataset & Feature: Olympic Sports, HMDB51 and UCF101; IDT + FV
- Semantic embedding: 300-d word vectors via the skip-gram model
- Visual-semantic mapping: a set of independent linear SVMs

> Quantitative results:

(1) ZSECOC vs. other embeddings

		•	``				
Olympic Spor	mpic Sports HMDB51		Method		Olympic Sports	HMDB51	UCF101
$\frac{216+09}{216+09}$	16.5+	-3.9 32+0.7	R	SECOC [3]	$18.4{\pm}0.6$	5.3 ±0.1	$3.0{\pm}0.8$
27.7 ± 4.6	N/2	A 13.7 ± 0.5	R	DECOC [3]	25.3 ± 1.8	6.2 ± 1.0	$2.7{\pm}0.2$
59.8±5.6	22.6	-1.2 15.1±1.7	D	DECOC [42]	40.1 ± 4.2	6.9 ± 1.0	$4.8 {\pm} 0.6$
			Fores	t-ECOC [9]	$51.0{\pm}8.7$	$9.2{\pm}0.7$	$5.9{\pm}0.5$
(3) 7SECOC ve state_of_the_art				ZSECOC	59.8±5.6	22.6±1.2	15.1±1.7
		IIC-ait					
Reference	Feature	Label Embedding	TD	Olympic Spor	ts HMDB51	J	JCF101
CVPR 2011	FV	А	Х	46.1±12.4	N/A	14.9±0.8	
TPAMI 2014	FV	А	\times	$45.4{\pm}12.8$	N/A	14.3 ± 1.9	
TPAMI 2014	FV	А	\times	42.3 ± 12.5	N/A	12	$2.8{\pm}2.0$
ICIP 2015	BoW	WV	\times	N/A	13.0 ± 2.7	1	0.9 ± 1.5
ICIP 2015	BoW	WV	\checkmark	N/A	15.0 ± 3.0	1	5.8±2.3
ICML 2015	FV	WV	\times	39.6 ± 9.6	18.5 ± 2.0	15.0 ± 1.3	
CVPR 2015	FV	WV	\times	28.6 ± 4.9	13.3 ± 2.4	9.9 ± 1.4	
CVPR 2015	FV	А	\times	47.5 ± 14.8	N/A	12.0 ± 1.2	
ICCV 2015	FV	А	\checkmark	N/A	N/A	13.2 ± 1.9	
ICCV 2015	FV	A+WV	×	N/A	N/A	14	$4.0{\pm}1.8$
ECCV 2016	FV	WV	Х	44.3±8.1	19.7±1.6	1	5.8±1.3
Ours	FV	ECOC	×	59.8±5.6 (40-b	it) 22.6 ± 1.2 (70-	-bit) $15.1\pm$	1.7 (100-bit)
	Olympic Spor 21.6 \pm 0.9 27.7 \pm 4.6 59.8 \pm 5.6 COC vs. st Reference CVPR 2011 TPAMI 2014 TPAMI 2014 ICIP 2015 ICIP 2015 ICIP 2015 ICIP 2015 ICIP 2015 ICVPR 2015 ICVPR 2015 ICVPR 2015 ICCV 2015 ICCV 2015 ICCV 2015 ICCV 2015 ICCV 2015 ICCV 2015	Olympic Sports HMD 21.6 ± 0.9 $16.5\pm$ 27.7 ± 4.6 N/ 59.8 ± 5.6 $22.6\pm$ COC vs. state-of-t Reference Feature CVPR 2011 FV TPAMI 2014 FV TPAMI 2014 FV ICIP 2015 BoW ICIP 2015 BoW ICML 2015 FV CVPR 2015 FV ICCV 2016 FV	Olympic Sports HMDB51 UCF101 21.6±0.9 16.5±3.9 3.2±0.7 27.7±4.6 N/A 13.7±0.5 59.8±5.6 22.6±1.2 15.1±1.7 COC vs. stete-of-the-art Isolation Reference Feature Label Embedding CVPR 2011 FV A TPAMI 2014 FV A ICIP 2015 BoW WV ICIP 2015 BoW WV ICML 2015 FV WV CVPR 2015 FV A ICIP 2015 FV A ICCV 2016 FV WV	Olympic SportsHMDB51UCF101 21.6 ± 0.9 16.5 ± 3.9 3.2 ± 0.7 R 27.7 ± 4.6 N/A 13.7 ± 0.5 R 59.8±5.622.6±1.215.1±1.7 CForesCOC vs. state-of-the-artCOC vs. state-of-the-artReferenceFeatureLabel EmbeddingTDCVPR 2011FVA×TPAMI 2014FVAXICIP 2015BoWWVICIP 2015BoWWV×ICIP 2015FVWV×ICML 2015FVA×ICCVPR 2015FVA×ICCV 2015FVA×ICCV 2015FVA×ICCV 2015FVA+WV×ECCV 2016FVWV×OursFVECCC×	Olympic SportsHMDB51UCF101Method 21.6 ± 0.9 16.5 ± 3.9 3.2 ± 0.7 RSECOC [3] 27.7 ± 4.6 N/A 13.7 ± 0.5 RDECOC [3] 59.8 ± 5.6 22.6 ± 1.2 15.1 ± 1.7 DECOC [42]Forest-ECOC [9]ZSECOCCOC vs. state-of-the-artTDOlympic SportCVPR 2011FVA \times 46.1 ± 12.4 TPAMI 2014FVA \times 45.4 ± 12.8 TPAMI 2014FVA \times 42.3 ± 12.5 ICIP 2015BoWWV \checkmark N/AICIP 2015BoWWV \checkmark N/AICML 2015FVWV \times 39.6 ± 9.6 CVPR 2015FVA \checkmark 47.5 ± 14.8 ICCV 2015FVA \checkmark 47.5 ± 14.8 ICCV 2015FVA \checkmark N/A ICCV 2015FVA \checkmark N/A ICCV 2015FVA+WV \checkmark N/A ICCV 2015FVA+WV \times N/A ICCV 2016FVWV \times 44.3 ± 8.1 OursFVECOC \times 59.8 ± 5.6 (40-b	Olympic Sports HMDB51 UCF101 Method Olympic Sports 21.6 \pm 0.9 16.5 \pm 3.9 3.2 \pm 0.7 RSECOC [3] 18.4 \pm 0.6 27.7 \pm 4.6 N/A 13.7 \pm 0.5 RDECOC [3] 25.3 \pm 1.8 59.8 \pm 5.6 22.6 \pm 1.2 15.1 \pm 1.7 DECOC [42] 40.1 \pm 4.2 Forest-ECOC [9] 51.0 \pm 8.7 DECOC [42] 40.1 \pm 4.2 Forest-ECOC [9] 51.0 \pm 8.7 DECOC [42] 40.1 \pm 4.2 CVPR 2011 FV A \times 46.1 \pm 12.4 N/A TPAMI 2014 FV A \times 45.4 \pm 12.8 N/A ICIP 2015 BoW WV \times N/A 13.0 \pm 2.7 ICIP 2015 BoW WV \times N/A 13.0 \pm 2.7 ICIP 2015 BoW WV \times N/A 13.0 \pm 2.7 ICIP 2015 BoW WV \times N/A 13.0 \pm 2.7 ICIP 2015 FV WV \times 39.6 \pm 9.6 18.5 \pm 2.0 CVPR 2015 FV A	Olympic Sports HMDB51 UCF101 Method Olympic Sports HMDB51 21.6 \pm 0.9 16.5 \pm 3.9 3.2 \pm 0.7 RSECOC [3] 18.4 \pm 0.6 5.3 \pm 0.1 27.7 \pm 4.6 N/A 13.7 \pm 0.5 RDECOC [3] 25.3 \pm 1.8 6.2 \pm 1.0 59.8 \pm 5.6 22.6 \pm 1.2 15.1 \pm 1.7 DECOC [42] 40.1 \pm 4.2 6.9 \pm 1.0 Forest-ECOC [9] 51.0 \pm 8.7 9.2 \pm 0.7 ZSECOC 59.8 \pm 5.6 22.6 \pm 1.2 15.1 \pm 1.7 COC vs. state-of-the-art TD Olympic Sports HMDB51 U Reference Feature Label Embedding TD Olympic Sports HMDB51 U CVPR 2011 FV A \times 46.1 \pm 12.4 N/A 14 TPAMI 2014 FV A \times 442.3 \pm 12.5 N/A 14 ICIP 2015 BoW WV \checkmark N/A 13.0 \pm 2.7 16 ICIP 2015 BoW WV \checkmark N/A 13.3 \pm 2.4 99 <

			•	X 7				
Embedding	Olympic Spor	ts HMD	B51 UCF101	Method Ol		Olympic Sports	HMDB51	UCF101
Word Vector	$\frac{216+09}{216+09}$	16 5+	-39 32+07	R	SECOC [3]	$18.4{\pm}0.6$	5.3 ±0.1	$3.0{\pm}0.8$
Attribute	27.7 ± 4.6	N/A	4 13.7+0.5	R	DECOC [3]	$25.3 {\pm} 1.8$	6.2 ± 1.0	$2.7{\pm}0.2$
ZSECOC	<u>59.8±5.6</u>	22.6 ±	$1.2 15.1 \pm 1.7$	D	ECOC [42]	$40.1 {\pm} 4.2$	$6.9{\pm}1.0$	$4.8 {\pm} 0.6$
				Fores	t-ECOC [9]	$51.0{\pm}8.7$	$9.2{\pm}0.7$	$5.9{\pm}0.5$
(3) 7SECOC ve state of the art				ZSECOC	59.8±5.6	22.6±1.2	15.1±1.7	
(3) ZSEC			ne-ait					,
Method	Reference	Feature	Label Embedding	TD	Olympic Spor	HMDB51	L	JCF101
HAA [28]	CVPR 2011	FV	А	×	46.1±12.4	N/A	14.9±0.8	
DAP [26]	TPAMI 2014	FV	А	\times	$45.4{\pm}12.8$	N/A	14.3 ± 1.9	
IAP [26]	TPAMI 2014	FV	А	\times	42.3 ± 12.5	N/A	12	$2.8{\pm}2.0$
ST [59]	ICIP 2015	BoW	WV	\times	N/A	13.0 ± 2.7	10.9 ± 1.5	
ST [59]	ICIP 2015	BoW	WV	\checkmark	N/A	15.0 ± 3.0	15.8±2.3	
ESZSL [46]	ICML 2015	FV	WV	\times	39.6 ± 9.6	18.5 ± 2.0	15.0 ± 1.3	
SJE [2]	CVPR 2015	FV	WV	\times	28.6 ± 4.9	13.3 ± 2.4	9.9 ± 1.4	
SJE [2]	CVPR 2015	FV	А	\times	47.5 ± 14.8	N/A	12.0 ± 1.2	
UDA [23]	ICCV 2015	FV	А	\checkmark	N/A	N/A	13.2 ± 1.9	
UDA [23]	ICCV 2015	FV	A+WV	\times	N/A	N/A	14	$4.0{\pm}1.8$
MTE [60]	ECCV 2016	FV	WV	Х	44.3±8.1	19.7±1.6	15.8±1.3	
ZSECOC	Ours	FV	ECOC	×	59.8±5.6 (40-b	it) 22.6 ± 1.2 (70-	-bit) $15.1\pm$	1.7 (100-bit)

Qualitative results:

(1) Similarity matrices



(3) Top-5 returned video examples for unseen categories



















(2) ZSECOC vs. other ECOC













