Attribute Learning for Understanding Unstructured Social Activity

Yanwei Fu, Timothy M. Hospedales, Tao Xiang, and Shaogang Gong

{yanwei.fu,tmh,txiang,sgg}@eecs.qmul.ac.uk

EECS, Queen Mary University of London, London E1 4NS, UK

Queen Mary University of London

1. Introduction

Problem: automatic classification and annotation of unstructured group social activity videos, shared by social media platforms (e.g. Youtube).

Challenges:

- **Unconstrained videos** have more complex events than commercial videos.
- **Consumer videos** are captured with low resolution, poor lighting, occlusion, clutter, camera shake and background noise.
- **User annotations are poor quality**: sparse, incomplete and ambiguous.

Solutions and contributions:

- **Semi-latent attribute space** is introduced; it spans manually specified human semantic annotations and latent data-driven attributes.
- **SLAS** smoothly bridges the gap between zero-shot and N-shot learning; and small and large attribute ontologies.
- **SLAS topic classification model** *jointly* learns user-defined (UD), class-conditional (CC) and background no-discriminative (BN) latent attributes; Jointly learning avoids CC and BN attribs. rediscovering UD attribs.
- **SLAS topic classification model** is robust to label noise in UD attribs, due to additionally leverage automatically discovered latent attributes for a more robust overall representation.
- **Zero-shot learning (ZSL):** We show for the first time how latent attributes can (surprisingly) be used to enhance ZSL.
- **ZSL on AwA [2]:** Our latent attribute framework allows us to exceed the ZSL performance of [2] on AwA dataset.
- **Unstructured social activity attribute (USAA)** is a new benchmark video dataset with 69 user-defined attributes, 8 classes and 1600 videos.

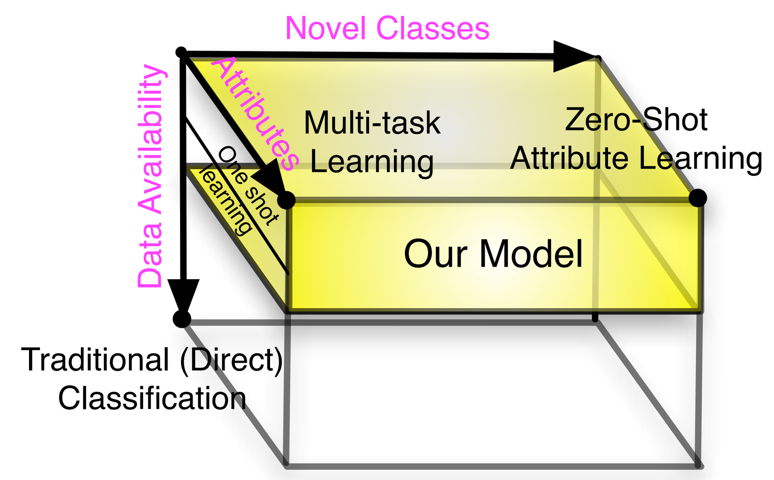
References

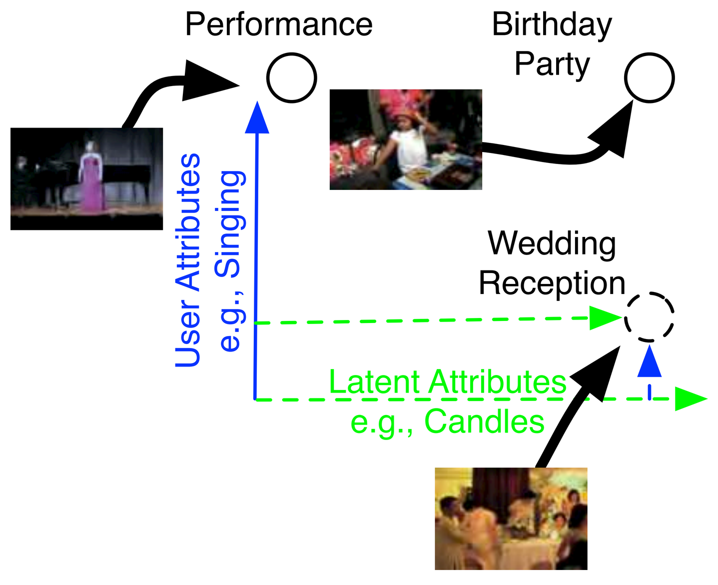
[1] Yu-Gang Jiang, Guangnan Ye, Shih-Fu Chang, Daniel Ellis, Alexander C. Loui. Consumer Video Understanding: A Benchmark Database and An Evaluation of Human and Machine Performance, ICMR 2011

[2] Christoph H. Lampert, Hannes Nickisch, Stefan Harmeling. Learning to detect unseen object classes by between-class attribute transfer. CVPR 2009.

2. Methodology

Semi-latent attribute space (SLAS) performs semantic feature reduction from the raw data to a lower dimensional semi-latent semantic attribute space.





Formalisation

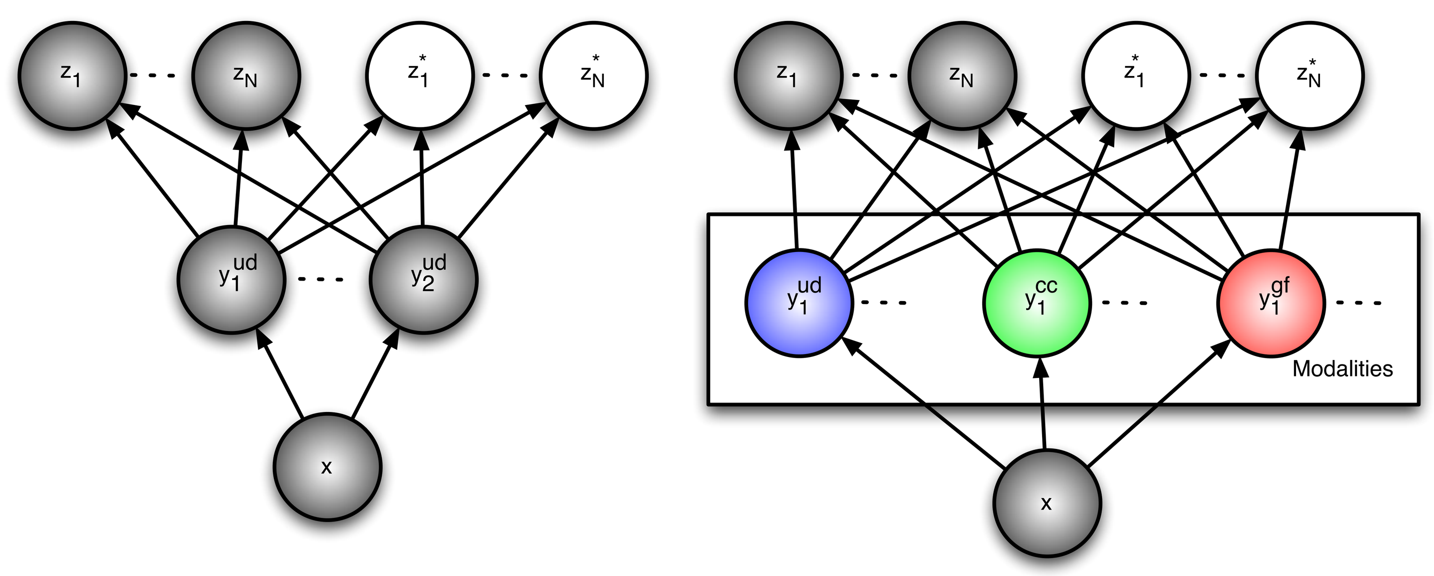
$$F = S(L(\cdot)),\ L:\mathcal{X}^d \rightarrow \mathcal{Y}^p,\ S:\mathcal{Y}^p \rightarrow \mathcal{Z}$$

L maps the raw data to semi-latent attribute space Y^p, and S further maps it to the final class Z.

ZSL with latent attributes:

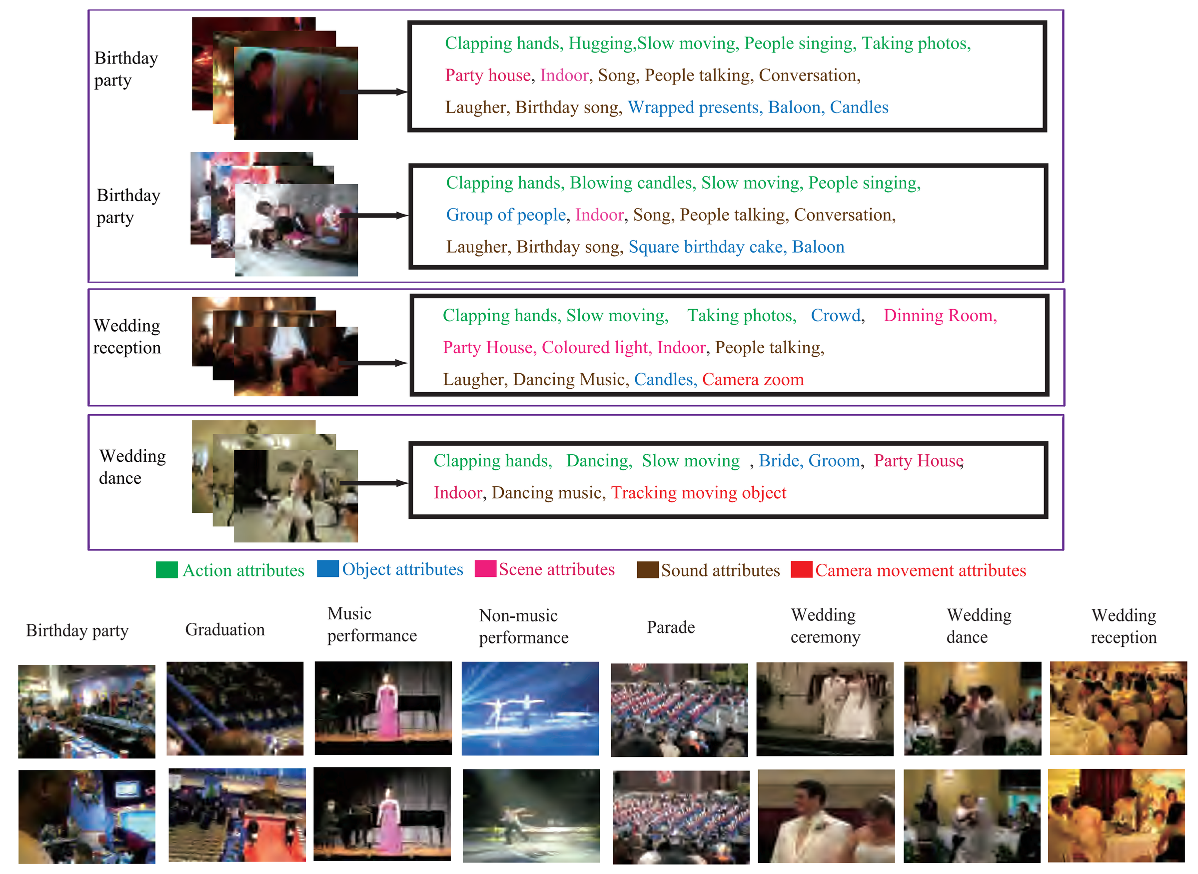
1. Input prototypes $\mathbf{v}_{z^*}^{ud}$ for novel classes z^*. Infer attributes γ^* for novel test data $\mathbf{x}^*$
2. NN matching of $\gamma^{*,ud}$ against prototypes $\mathbf{v}_{z^*}^{ud}$ in user-defined attribute space $\mathcal{Y}_{ud}$.
3. For each novel class z^*: (1) Find top-K most confident test-set matches $\{\gamma_{l,z^*}\}_{l=1}^{K}$; (2) Self train a new prototype in the full attribute-space: $\mathbf{v}_{z^*}=\frac{1}{K}\sum_l \gamma_{l,z^*}$.
4. NN matching of γ^* against prototypes $\mathbf{v}_{z^*}$ in full attribute space $\mathcal{Y}$.

DAP[2] vs ZSL with latent attributes:



3. Experiments

Example attributes and frames in USAA dataset: Manually annotated groundtruth attributes for 8 semantic class videos of CCV dataset[1] (100 videos per-class for training and testing).



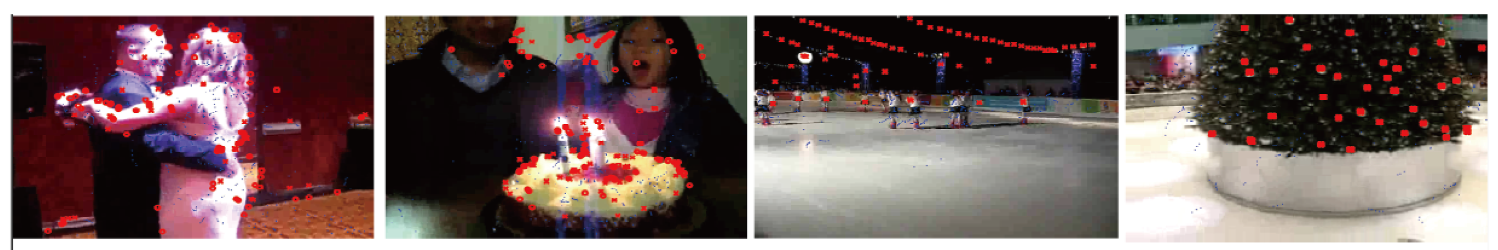
Classification: We extract SIFT, STIP and MFCC features [1]. SLAS can greatly improve the performance of transfer learning and ZSL. 'UD' refers to the number of user-defined attributes. N-shot transfer learning performance:

	1-shot			5-shot		
UD	KNN	SVM	Ours	KNN	SVM	Ours
0	30	–	**34**	34	–	**42**
7	30	32	**33**	34	43	**44**
34	30	**37**	35	34	47	**48**

Zero-shot learning classification. We compare DAP [2] vs. SLAS. "NF" refers to learn our model only with UD and CC attributes.

Continuous,0-shot			Binary,0-shot		
UD	UD+Latent		UD	UD+Latent	
DAP	Ours	NF	DAP	Ours	NF
38	**45**	41	31	**36**	31

Visualization of learned attributes: Corresponding interest-points are drawn with red points; blue dots indicate background points.



Visualization of user defined attributes

Visualization of latent attributes