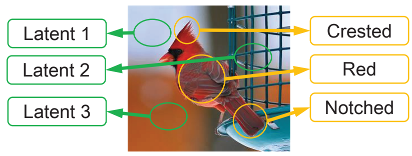
Joint Learning of Semantic and Latent Attributes

Peixi Peng[1], Yonghong Tian[1], Tao Xiang[2], Yaowei Wang[3], Tiejun Huang[1]

[1]National Engineering Laboratory for Video Technology, Peking University, China
[2]School of EECS, Queen Mary University of London, UK
[3]Department of Electronic Engineering, Beijing Institute of Technology, China
{pxpeng, yhtian, tjhuang}@pku.edu.cn, yaoweiwang@bit.edu.cn, t.xiang@qmul.ac.uk

1 Motivations

- Attributes are effective mid-level representations for cross-class transfer learning problems such as zero-shot learning (ZSL) and person re-identification (Re-ID).
- Joint learning of semantic and latent attributes (both discriminative and non-discriminative/background ones) are important:
  - More comprehensive discriminative representations
  - Better attributes prediction with background attributes

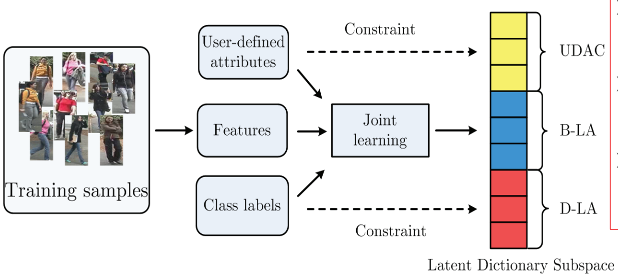


Avoid the unrelated parts corrupting the learned semantic attribute predictor.

- Contributions
  - A unified framework for learning both user-defined semantic attributes and discriminative latent attributes is proposed.
  - We further develop a novel dictionary learning model which decomposes the learned dictionary subspace into three parts corresponding to the semantic, discriminative latent as well as background latent attributes respectively.

2 Methodology

Decomposition of the dictionary:
- UDAC (D^u): correlated to the semantic attributes
- D-LA (D^d): the discriminative latent attributes correlated to the class labels
- B-LA (D^b): the background latent attributes which capture all the residual information



The stepwise form:
- The minimization of the first reconstruction error term enables UDAC and D-LA to encode the feature as much as possible.
- The minimization of the second reconstruction error term enables B-LA to encode and align the residual part.
- The combination of these two terms is to prevent the B-LA from dominating the reconstruction error leading to trivial solutions for UDAC and D-LA.

$$[D^u, D^d, D^b, W] = \arg\min \|Y - D^u X^u - D^d X^d\|_F^2 + \|Y - D^u X^u - D^d X^d - D^b X^b\|_F^2$$
$$+ \beta^2 \|X^u - WA\|_F^2 + \alpha \sum_{i,j=1}^{n} m_{i,j} \|x_i^d - x_j^d\|^2, \quad s.t. \|d_i^u\|_2^2 \le 1, \|d_i^d\|_2^2 \le 1, \|d_i^b\|_2^2 \le 1, \|w_i\|_2^2 \le 1.$$

The constraint for UDAC:
- To stablish a linear constraint W between the projection in UDAC and semantic attribute annotations A
- The reason is that the sematic attributes cannot act as suitable bases for a subspace.

The constraint for D-LA
- The graph Laplacian regularization to dictate that the D-LA subspace to be more discriminative

The regularization term:
- To avoid overfitting

Algorithm 1: The proposed algorithm

Input: X_t; initialise D^u, D^d, D^b and W randomly; $X^d \rightarrow \mathbf{0}$, $X^b \rightarrow \mathbf{0}$;
Output: D^u, D^d, D^b, X^u, X^d, X^b and W.
while *Non-convergence* **do**
  Coding problem:
    compute code X^u.
    compute code X^d.
    compute code X^b.
  Updating dictionaries:
    update D^u and D^d.
    update D^b.
  Updating W:
    update W .

3 Experiments

- Zero-shot Learning and Attribute Prediction (ATT)

	ZSL		ATT	
DATASET	AwA	CUB	AwA	CUB
DAP	57.5	-	72.8	61.8
ALE	43.5	18.0	65.7	60.3
UMF	48.6	18.2	-	-
CSHAP	45.6	17.5	**74.3**	68.7
SSE	76.3	30.4	-	-
SJE	61.9	40.3	-	-
JLSE	80.5	42.1	-	-
LatEm	76.1	51.7	-	-
Ours	**82.1**	**56.5**	73.6	**78.3**

The accuracy and mAUC are reported for ZSL and ATT respectively. There is a performance gain compared to the published version of this paper since we optimize the implementation method. For more details see the upcoming open codes.

- Person Re-ID

	VIPeR	PRID	Market	iLIDS
RPLM	27.0	15.0	-	-
kCCA	37.0	15.0	-	-
MFA	39.6	20.9	37.6	49.2
kLFDA	39.9	21.6	38.4	48.4
XQDA	40.0	-	-	-
MLAPG	40.7	-	-	-
Ours_L	**41.3**	**26.8**	**47.4**	**52.3**
aMTL	42.3	18.0	-	-
Ours_U	28.4	16.3	-	39.2
Ours_All	**45.4**	**26.8**	**-**	**56.8**

The rank 1 are reported. Ours_L means only D-LA are used, Ours_U means only semantic attributes are used, and Ours_All means both type of attributes are used for person Re-ID.

- Visualization of the learned D-LA

Most are visually meaningful in a subtle way and thus may have been ignored by human annotators, such as clothes with stripes, white logo on the chest, white coat with dark clothes and female with long hairs and short sleeve.