Zero-Shot Learning via Low-Rank-Representation Based Manifold Regularization

Min Meng, *Member, IEEE*, and Xiaoyu Zhan

***Abstract*—In this letter, we propose a novel low-rank-representation (LRR) based manifold-regularization approach for zero-shot learning (ZSL). Most existing regularization-based ZSL approaches perform the alignment between visual feature space and semantic space based on the affinity matrix constructed from the test instances. The affinity matrix plays a significant role in exploiting the manifold structures of visual feature space, hence we propose to use the LRR to guide the affinity-matrix construction by exploring the subspace structures of data. Considering the locality and similarity information among data, we incorporate a Laplacian regularization term to the LRR framework to ensure that the learned affinity matrix can capture the local geometric structures in data. We also explicitly impose the nonnegative sparse constraint on the affinity matrix to facilitate the learning of local manifold structures. Moreover, we use an effective manifold-regularization methodology to learn discriminative semantic representations of test instances, leading to significant improvements in classification performance over the unseen classes. Extensive experiments on three benchmark datasets demonstrate that the proposed approach outperforms the state of the arts.**

***Index Terms*—Affinity matrix, low-rank representation (LRR), manifold regularization, zero-shot learning (ZSL).**

I. Introduction

ZERO-SHOT learning (ZSL) for visual recognition aims at recognizing unseen classes that have no labeled samples for training, by transferring supervision information from related seen classes with abundant labeled samples. Many efforts have been made to fulfill this goal by incorporating an intermediate representation [1]–[11] to facilitate the knowledge transfer across class categories. In practical, researchers favorably utilize some semantic properties, such as attribute vectors [12]–[16] or word vectors [8], [17]–[22], acting as the intermediate representation to establish the interclass connections. How to effectively encode the relationship between seen and unseen classes with the semantic representation is a key problem for ZSL tasks. However, most existing methods generally construct the semantic embedding without considering the underlying information in visual feature space, which makes it not correlated to the visual features. Besides, such representations could have uncertainty or inconsistencies of their own, e.g., if they are obtained by some measurement device.

Recently, some approaches have been proposed to address the problem from the perspective of manifold learning and shown the appealing superiority for ZSL. Specifically, many regularization-based methods have been applied to ZSL to improve the performance of classification [23]–[26]. For instance, Fu *et al.* [23] proposed a multiview transductive approach through aligning multiview data using canonical correlation analysis. Zhang *et al.* [24] developed a structured-prediction method based on test-time adaptation of similarity functions learned using training data. More recently, Li *et al.* [25] exploited the intrinsic relationship between the semantic space and the transfer ability of visual semantic mapping to generate optimized semantic embedding space. Deutsch *et al.* [26] proposed an alignment method using a multiscale graph transform, which can perform the alignment locally while taking into account the global properties of the space. In addition, there are some other works sharing the similar idea for joint alignment of the data using sparse representations [27], [28] or transfer learning [29]–[31]. For instance, Kodirov *et al.* [27] proposed a regularized sparse-coding model to deal with the domain shift problem in ZSL. Qiao *et al.* [28] proposed a denoising-based alignment method by introducing an $l_{2,1}$ objective function as the noise-suppression regularizer.

A. Motivation

Although some attempts have already been made, the problem of encoding the relationships between classes in the semantic space has not been investigated well. Most existing regularization-based ZSL methods [23]–[26] perform the alignment of two different manifolds in visual feature space and semantic space based on the affinity matrix. Therefore, the affinity matrix plays a significant role in exploiting the manifold structure of visual feature space to help obtain a better semantic embedding. However, these methods construct the affinity matrix locally by using some metrics between the observations, e.g., the cosine similarity, and neglect the intrinsic manifold structures of the data. Thus, it is necessary to utilize the advantages of subspace learning to guide the affinity matrix, such that it can deliver strong discriminant information. During past decades, a number of robust subspace learning methods [32]–[34] have been proposed to recover such subspace structure. Given a set of instances, which may be corrupted by errors and approximately drawn from the subspaces, we explore the subspace structures by a low-rank representation (LRR) model to learn an affinity matrix with strong discriminative ability. Moreover, we also incorporate a Laplacian regularization term into the LRR model such that both the global mixture of subspace structure and the

Manuscript received March 30, 2018; revised May 10, 2018; accepted July 14, 2018. Date of publication July 18, 2018; date of current version August 3, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61702114 and Grant 61672171 and in part by the Foundation for Young Innovative Talents in Higher Education of Guangdong, China under Grant 2016KQNCX040. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mario Huemer. (*Corresponding author: Min Meng.)*

The authors are with the Department of Computer Science, Guangdong University of Technology, Guangzhou 510006, China (e-mail: mengmin1985@gmail.com; zhanxyz@outlook.com).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2018.2857201

local geometric structure of data can be exploited. In addition, we also explicitly impose the nonnegative sparse constraint on the affinity matrix to facilitate learning local manifold structures in our framework.

After obtaining the affinity matrix, we propose to use a manifold-regularization methodology to exploit the semantic vectors of test instances. Despite performing regularization like in [35], the advantages of our approach are threefold compared against [35]. First, our approach can exploit the subspace structures and locality information of data so that the learned affinity matrix is optimal to deliver the strong discriminative information. Moreover, the proposed method can correct the possible errors and simultaneously learn the discriminative affinity matrix benefiting from our LRR framework. Finally, the proposed LRR model is general and can benefit other regularization-based approaches, such as in [23]–[26]. We will show that the affinity matrix learned does provide an effective way of encoding both the global low-dimensional structures and the local geometric structures of data to help distinguish categories, leading to significant performance improvements.

Our key contributions are summarized as follows.

1) We model the manifold structures of the visual feature space by a LRR model to explore the subspace structures. Besides, a Laplacian regularization term is incorporated to guide the affinity matrix such that we can effectively exploit both the global subspaces structure and the local geometric structure of data.
2) To facilitate learning local manifold structures, we explicitly impose the nonnegative sparse constraints on the affinity matrix. As we intend to use the locally linear subspaces to approximate the nonlinear manifold, the nonnegative sparse constraint leads to a sparse low-rank matrix that represents each data instance as a linear combination of others.
3) We employ a manifold-regularization technique to learn semantic vectors of test instances using regularization adapted to the manifold domain. Owing to the strong discrimination delivered by the learned affinity matrix, our approach can achieve significant recognition performance on unseen classes compared to previous works.

II. Methodology

The training set consists of samples $S = \{(\mathbf{x}_s^i, \mathbf{y}_s^i, z_s^i)\}_{i=1}^{n_s}$, where $\mathbf{x}_s^i \in X$, $X \subset \mathbb{R}^d$ denotes the visual space, $\mathbf{y}_s^i \in Y$, $Y \subset \mathbb{R}^m$ is the semantic or attribute space, and $z_s^i \in \{0,1\}^{c_s}$ indicates an observed label in a set of cardinality c_s. We are interested in classifying the test set $\{\mathbf{x}_t^j\}_{j=1}^{n_t}$ into a different set of c_t unseen classes. Assuming that the training data are noisy values assigned at samples around the manifold X, we formulate the problem as the alignment of a smooth Y-valued function on X.

A. Learning LRR-Based Affinity Matrix

Assume the data X is approximately drawn from a union of k low-dimensional subspace $\{\prod\}_{i=1}^k$ contaminated by error E, the objective function of LRR model can be formulated as

$$\min_{Z,E} \quad \operatorname{rank}(Z) + \gamma\|E\|_0 \quad \text{s.t.} \quad X = XZ + E \tag{1}$$

where γ is a parameter; the data X is used as the dictionary; $\|\cdot\|_0$ is the l_0 pseudonorm, which is defined as the number of nonzero entries; and $Z = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{n_s}]$ is the coefficient matrix with each $\mathbf{z}_i$ being the representation of $\mathbf{x}_i$. Due to the discrete nature of the rank function and the l_0 pseudonorm, direct optimization of (1) is NP-hard. Thus, the problem of (1) is relaxed into the optimization problem as follows:

$$\min_{Z,E} \quad \|Z\|_* + \gamma\|E\|_1 \quad \text{s.t.} \quad X = XZ + E \tag{2}$$

where $\|Z\|_*$ is the nuclear norm of Z, which can approximate the rank of Z; and $\|E\|_1$ is a good relaxation of the l_0 pseudonorm of E. As the data X is used as the dictionary directly, then the element z_{ij} of Z reflects the similarity between data pairs $\mathbf{x}_i$ and $\mathbf{x}_j$; hence, Z can be used as the affinity matrix.

To introduce richer information over Z, some helpful regularizations can be imposed on the coefficient matrix. We intend to use locally linear subspaces to approximate the nonlinear manifold. Meanwhile, the matrix Z learned from (2) needs to be pruned to obtain the optimal affinity matrix. To this end, we impose a nonnegative sparse constraint $Z \geq 0$ and $\|Z\|_0 \leq T$ to ensure that the obtained low-rank and sparse matrix can be directly used as the affinity matrix that represents each instance as a linear combination of others. Thus, the nonnegative sparse LRR model can be formulated as follows:

$$\begin{aligned} &\min_{Z,E} \quad \|Z\|_* + \gamma\|E\|_1 \\ &\text{s.t.} \quad X = XZ + E, Z \geq 0, \|Z\|_0 \leq T. \end{aligned} \tag{3}$$

Inspired by the graph-based manifold learning [36], we also incorporate the Laplacian regularization term $\sum_{i,j} W_{i,j}\|\mathbf{z}_i - \mathbf{z}_j\|^2$ into the optimization problem of (3), where $\mathbf{z}_i$ and $\mathbf{z}_j$ are the new representations of $\mathbf{x}_i$ and $\mathbf{x}_j$, respectively; and the weight W_{ij} is set as 1 if $\mathbf{x}_i \in N_k(\mathbf{x}_j)$ or $\mathbf{x}_j \in N_k(\mathbf{x}_i)$. Under this formulation, we can assume that if $\mathbf{x}_i$ and $\mathbf{x}_j$ are close in the data space, then $\mathbf{z}_i$ and $\mathbf{z}_j$ should be as close as possible. Therefore, this regularization term can preserve the local geometric structure information characterized by the graph and help us learn a more compact and discriminative coefficient matrix. Thus, the nonnegative sparse Laplacian regularized LRR model is formulated as follows:

$$\begin{aligned} &\min_{Z,E} \|Z\|_* + \gamma\|E\|_1 + \beta\sum_{i,j} W_{i,j}\|\mathbf{z}_i - \mathbf{z}_j\|^2 \\ &\text{s.t.} \quad X = XZ + E, Z \geq 0, \|Z\|_0 \leq T. \end{aligned} \tag{4}$$

The problem (4) can be solved by convex relaxation with the breakthroughs in high-dimensional optimization [37]. We employ the linearized alternating direction method with adaptive penalty [38] to solve the convex optimization associated with (4). Please refer to the details about the convex relaxation and optimization of (4) in the supplementary document. Instead of explicitly designing the coefficients of the affinity matrix, we choose $\frac{1}{2}(|Z| + |Z^T|)$ as the symmetrized affinity matrix, which is learned from the data X.

B. Manifold Regularization

Here, we focus on addressing the manifold regularization problem by exploring graph-based equivalents of classical denoising algorithms [39]. In this letter, we use the semantic attribute as the graph signal. Specifically, test instances are represented as an attributed graph, within which visual data $\mathbf{x}_i \in X_u$ are the nodes and the attributes $\mathbf{y}_i$ represent the graph signals.

For an arbitrary dimension r of the semantic attributes $\mathbf{y}$, we have a noisy graph signal $\tilde{\mathbf{f}}_r = \mathbf{f}_r + \eta_r$, where η_r is

Algorithm 1: The Proposed Regularization Approach.

Input: Unlabeled instances X_u and their semantic representations $\mathbf{y}$.
Calculate Z by solving (4) using the visual data X_u;
Compute the affinity matrix $A = \frac{1}{2}(|Z| + |Z^T|)$;
Construct the Laplacian L from A.
for each dimension $r \in 1, \ldots, m$ **do**
1. Assign the corresponding coordinate values of $\mathbf{y}$ as the graph signal $\tilde{\mathbf{f}}_r$;
2. Calculate the regularized signal $\mathbf{f}_r$ by solving (6) with respect to L.
end
Output: Regularized semantic representations $\widehat{\mathbf{y}} = [\mathbf{f}_1, \ldots, \mathbf{f}_m]$.

uncorrelated noise and we aim to recover $\mathbf{f}_r$. Thus, we have that $\mathbf{y}_i = [\tilde{f}_1(i), \tilde{f}_2(i), \ldots, \tilde{f}_m(i)]$ for the instance $\mathbf{x}_i$. In order to characterize the global smoothness of the clean signal $\mathbf{f}_r$ with respect to the underlying graph, we define its graph Laplacian quadratic form as follows:

$$\|\nabla \mathbf{f}_r\|^2 = \sum_{i,j} A_{ij}(f_r(i) - f_r(j))^2 = \mathbf{f}_r^{\mathrm{T}} L \mathbf{f}_r \tag{5}$$

where L indicates the combinational graph Laplacian defined on the affinity matrix A as $L = D - A$, with D as the diagonal matrix whose diagonal entries are $d_{ii} = \sum_j A_{ij}$. Here $\frac{1}{2}(|Z| + |Z^T|)$ is used as the affinity matrix that is expected to correct the possible errors and simultaneously capture the global mixture of subspace structure and locally linear structure of the data X_u.

Similar to [35], we apply the Tikhonov regularization using the above quadratic form $\mathbf{f}_r^{\mathrm{T}} L \mathbf{f}_r$ for a fixed $\lambda > 0$ to recover $\mathbf{f}_r$. Then, the graph denoising corresponds to solve the optimization problem $\arg\min_{\mathbf{f}_r} \|\mathbf{f}_r - \tilde{\mathbf{f}}_r\|_2^2 + \lambda \mathbf{f}_r^{\mathrm{T}} L \mathbf{f}_r$. The solution can be obtained equivalently by solving

$$\mathbf{f}_r = (I + \lambda L)^{-1} \tilde{\mathbf{f}}_r. \tag{6}$$

Thus, we can obtain the denoised signal $\mathbf{f}_r$ by smoothing out the noisy signal $\tilde{\mathbf{f}}_r$ based on the underlying graph connectivity. As the manifold structure of the data X_u is encoded in the affinity matrix, the proposed regularization method allows us to perform the alignment to learn discriminative semantic representations of test instances. Suppose $\mathbf{y} \in \mathbb{R}^m$ denote the predicted semantic representations for test instances, we apply the proposed regularization algorithm to all semantic representation dimensions, to obtain the full regularized semantic representations $\widehat{\mathbf{y}} = [f_1, f_2, \ldots, f_m]$. The proposed regularization approach for ZSL is summarized in Algorithm 1.

C. *Zero-Shot Recognition*

Given the regularized semantic representation $\widehat{\mathbf{y}}_i$ for each instance $\mathbf{x}_i$ in the testing set, we can use the standard clustering method to globally partition the regularized graph constructed from $\widehat{\mathbf{y}}_i$ into c_t classes. In this letter, we adopt the spectral clustering [40] to perform test instances classification over the unseen classes.

III. Experiments

We test our method with predefined attributes for ZSL on the three benchmark datasets AwA, CUB, and aPY. For fair comparison with the state of the arts, we adopt the 4096-dim CNN features extracted from the verydeep-19 network [41]. In addition to human annotated attributes, word vector is exploited for AwA and CUB. We use the pre-extracted 400-dim word2vec vectors from wikipedia as in [19]. In practice, we use existing methods to obtain an initial mapping from the image feature space to the semantic attribute space on the training set. Then, the mapping is used to provide an initial estimation for test attributes based on input image features. In our experiments, we use the similar experimental settings as [24], including the CNN features and the initial estimation produced by [19]. Our method has three hyperparameters, the weights β and γ and the regularization parameter λ. We will conduct experiments to evaluate their sensitivity. All reported results are per-class top-1 accuracies averaged over ten trials in our experiments.

TABLE I
Comparison With State-of-the-art Methods on the Three Benchmark Datasets. In each column w.r.t attribute (A) or word vector (V), the results shown in boldface are better than the others

Methods	Semantic	AwA	CUB	aPY
DAP [1]	A	57.23	36.7	38.16
ESZSL [2]	A	75.32	39.04	24.22
MLZSC [20]	A	77.36	43.3	53.2
DeViSE [3]	A/W	56.7/50.4	33.5	–
Ba et al. [18]	A/W	69.3/58.7	34.0	–
SJE [19]	A+W	73.5	51	–
Ding et al. [11]	A	82.8	45.2	55.2
SSE [7]	A	76.33	30.41	46.23
JLSE [21]	A	80.46	42.11	50.35
UDAZSL [27]*	A+W	75.6	40.6	–
TMV-HLP [23]*	A	73.5	47.9	–
SP-ZSR [24]*	A	**92.06**	53.26	69.74
Li et al. [25]*	A	78.71	51.59	–
Qiao et al. [28]*	W	66.46	29.0	–
Deutsch et al. [26]*	W	80	35	–
Ours*	A/W	85.29/**82.39**	**53.26/49.64**	**85.22**

(∗ indicates the transductive methods.)

A. *Results and Discussion*

We select the 15 most recent and competitive ZSL methods for comparison on three benchmark datasets, as shown in Table I. Our overall results are superior to the compared methods on the benchmarks. Despite the lack of human annotation for attributes (Word2Vec), our approach still yields competitive results as compared with the state of the arts. This demonstrates the high discrimination delivered by learned affinity matrix and the effectiveness of the proposed regularization method. Among the compared methods, six of them and the proposed model are transductive methods, which aim to rectify the projection domain shift problem by utilizing the unlabeled test instances. As expected, these methods are very competitive and our method still yields superior performance. As opposed to the methods [23], [27], [28], which directly seek to associate test-data distribution with training data, the methods in [24]–[26] and ours accounting for test-time data shifts are more effective. Note that [24] obtained better result on AwA (92.06% versus 85.29%) but its results on aPY (69.74% versus 85.22%) are much weaker than that of ours. This is because [24] only explores local geometric structures of data to predict smooth labeling structures. Thus, it cannot work well if the initial predicted similarities

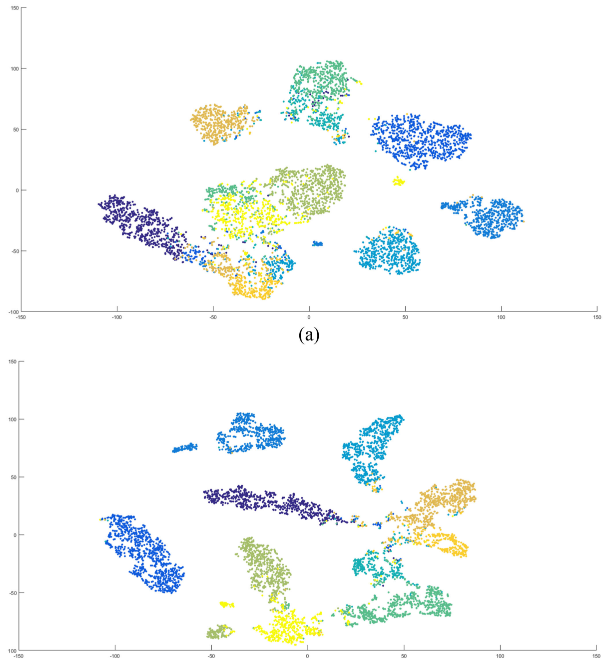

Fig. 1. tSNE visualization of the distribution of 10 unseen classes' images in the AwA dataset with attributes before and after performing regularization.

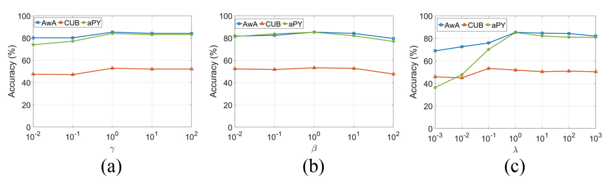

Fig. 2. Recognition accuracies on three datasets with attributes versus different parameter values: (a) $\gamma.(b)\beta$. (c)λ.

and empirical distributions of unlabeled instances are not distinguishing, like on aPY. In contrast, our method significantly outperforms related methods by more than 15% on aPY. The results validate that when the distribution of test instances are not well separable, our method is more effective due to the strong discrimination delivered by our learned affinity matrix, which can effectively exploit the global mixture of subspace structure as well as the local geometric structure of data.

To verify the effectiveness of the proposed regularization method, we plot the distribution of 10 unseen classes' images in AwA with semantic attributes before and after performing regularization, as shown in Fig. 1. The result in Fig. 1(a) indicates that it is challenging to estimate the data distribution for unseen classes using the initial noisy semantic vectors. The result in Fig. 1(b) validates that our approach is capable of generating high-quality estimation results through learning new semantic vectors from noisy instances using regularization applied to the manifold domain.

We further study the influence of the parameters γ, β, and λ. We vary a parameter while keeping others fixed. The results with different parameter values on three datasets are given in Fig. 2. The performance of the parameter's variations are similar on three datasets. We can see that the performance of our method is insensitive to the choice of γ and β. In particular, λ is used to control the weight of manifold-regularization term. We can

TABLE II
RECOGNITION ACCURACY (%) OF THE PROPOSED REGULARIZATION METHOD USING DIFFERENT VARIANTS OF AFFINITY MATRIX. IN EACH COLUMN W.R.T ATTRIBUTE (A) OR WORD VECTOR (V), THE RESULTS SHOWN IN BOLDFACE ARE BETTER THAN THE OTHERS

Methods	AwA		CUB		aPY
	att	w2v	att	w2v	att
Regularization using A_{NNG}	81.59	66.36	49.11	44.99	81.14
Regularization using A_{LRR}	82.27	70.02	50.13	46.51	83.27
Proposed	**85.29**	**82.39**	**53.26**	**49.64**	**85.22**

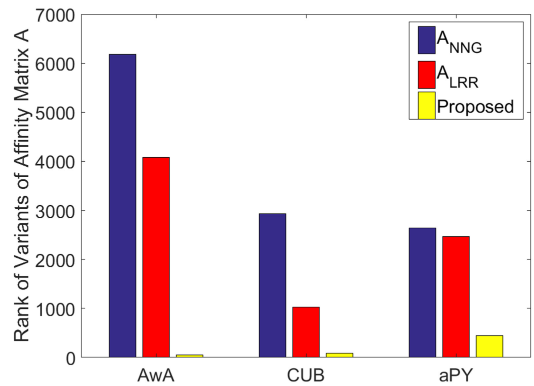

Fig. 3. The rank of different variants of affinity matrix for unseen classes on two datasets.

also note that the performance of our method is also robust to parameter λ when $\lambda \geq 1$. In our experiments, we set $\beta = 5$, $\gamma = 1$, and $\lambda = 1$.

Finally, we validate the effectiveness of one key element of our method, i.e., affinity-matrix learning. To this end, two variants of affinity matrix are incorporated into our framework and the comparison results are reported in Table II. "A_{NNG}" represents the affinity matrix constructed from a k-nearest neighbor graph with weights defined using the cosine similarity between instances, and "A_{LRR}" indicates the affinity matrix constructed from the LRR model in (1). In this case, "Proposed" is equivalent to regularization using the affinity matrix constructed from the nonnegative sparse Laplacian regularized LRR model in (4). As can be seen, the proposed method achieves the highest accuracies on three datasets, validating the effectiveness of the affinity matrix chosen in our framework. We can conclude that our learned affinity matrix is more discriminative than "A_{NNG}" and "A_{LRR}," since it can effectively capture the global low-dimensional structures and the local geometric information of data. Fig. 3 shows the rank of different variants of affinity matrix for unseen classes on three datasets. The rank of affinity matrix learned in our method is the lowest, clearly demonstrating its advantage at delivering strong discriminative information over the others.

IV. CONCLUSION

In this letter, we addressed ZSL using an effective manifold-regularization framework based on LRR. We learned the affinity matrix with strong discriminant ability using the proposed LRR model, and further performed regularization to fully exploit semantic representations of test instances. Experimental results clearly show the superiority of our approach against the state of the arts.

References

[1] C. H. Lampert, H. Nickisch, and S. Harmeling, "Attribute-based classification for zero-shot visual object categorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 453–465, Mar. 2014.

[2] B. Romera-Paredes and P. H. S. Torr, "An embarrassingly simple approach to zero-shot learning," in *Proc. Int. Conf. Int. Mach. Learning*, 2015, pp. 2152–2161.

[3] A. Frome *et al.*, "Devise: A deep visual-semantic embedding model," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2013, pp. 2121–2129.

[4] M. Norouzi *et al.*, "Zero-shot learning by convex combination of semantic embeddings," in *Proc. Int. Conf. Learn. Representations*, 2014, pp. 1–9.

[5] R. Socher, M. Ganjoo, H. Sridhar, O. Bastani, C. D. Manning, and A. Y. Ng, "Zero-shot learning through cross-modal transfer," in *Proc. Conf. Neural Inf. Process. Syst.*, 2013, pp. 935–943.

[6] X. Wang and Q. Ji, "A unified probabilistic approach modeling relationships between attributes and objects," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 2120–2127.

[7] Z. Zhang and V. Saligrama, "Zero-shot learning via joint latent similarity embedding," in *Proc. Comput. Vis. Pattern Recognit.*, 2016, pp. 6034–6042.

[8] Y. Xian, Z. Akata, G. Sharma, Q. Nguyen, M. Hein, and B. Schiele, "Latent embeddings for zero-shot classification," in *Proc. Comput. Vis. Pattern Recognit.*, 2016, pp. 69–77.

[9] D. Wang, Y. Li, Y. Lin, and Y. Zhuang, "Relational knowledge transfer for zero-shot learning," in *Proc. AAAI Conf. Artif. Intell.*, 2016, pp. 2145–2151.

[10] J. Qin, Y. Wang, L. Liu, J. Chen, and L. Shao, "Beyond semantic attributes: Discrete latent attributes learning for zero-shot recognition," *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1667–1671, Nov. 2016.

[11] Z. Ding, M. Shao, and Y. Fu, "Low-rank embedded ensemble semantic dictionary for zero-shot learning," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, 2017, pp. 6005–6013.

[12] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth, "Describing objects by their attributes," in *Proc. IEEE Trans. Comput. Vis. Pattern Recognit.*, 2009, pp. 1778–1785.

[13] C. H. Lampert, H. Nickisch, and S. Harmeling, "Learning to detect unseen object classes by between-class attribute transfer," in *Proc. IEEE Comput. Visi. Pattern Recognit.*, 2009, pp. 951–958.

[14] D. Parikh and K. Grauman, "Relative attributes," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 503–510.

[15] W. Wang, Y. Yan, S. Winkler, and N. Sebe, "Category specific dictionary learning for attribute specific feature selection," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1465–1478, Mar. 2016.

[16] Y. Guo, G. Ding, J. Han, and Y. Gao, "Zero-shot learning with transferred samples," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3277–3290, Jul. 2017.

[17] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *Adv. Neural Inf. Process. Syst.*, vol. 26, pp. 3111–3119, 2013.

[18] J. L. Ba, K. Swersky, S. Fidler, and R. Salakhutdinov, "Predicting deep zero-shot convolutional neural networks using textual descriptions," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 4247–4255.

[19] Z. Akata, S. Reed, D. Walter, H. Lee, and B. Schiele, "Evaluation of output embeddings for fine-grained image classification," in *Proc. Comput. Vis. Pattern Recognit.*, 2015, pp. 2927–2936.

[20] M. Bucher, S. Herbin, and F. Jurie, "Improving semantic embedding consistency by metric learning for zero-shot classiffication," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 730–746.

[21] Z. Zhang and V. Saligrama, "Zero-shot learning via semantic similarity embedding," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 4166–4174.

[22] L. S. F. S. G. D. Yang Long, Li Liu, and J. Han, "From zero-shot learning to conventional supervised classification: Unseen visual data synthesis," in *Proc. Comput. Vis. Pattern Recognit.*, 2017, pp. 6165–6174.

[23] Y. Fu, T. M. Hospedales, T. Xiang, and S. Gong, "Transductive multi-view zero-shot learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 11, pp. 2332–2345, Nov. 2015.

[24] Z. Zhang and V. Saligrama, "Zero-shot recognition via structured prediction," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 533–548.

[25] Y. Li, D. Wang, H. Hu, Y. Lin, and Y. Zhuang, "Zero-shot recognition using dual visual-semantic mapping paths," in *Proc. Comput. Vision Pattern Recognit.*, 2017, pp. 5207–5215.

[26] S. Deutsch *et al.*, "Zero shot learning via multi-scale manifold regularization," in *Proc. Comput. Vis. Pattern Recognit.*, 2017, pp. 5292–5299.

[27] E. Kodirov, T. Xiang, Z. Fu, and S. Gong, "Unsupervised domain adaptation for zero-shot learning," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 2452–2460.

[28] R. Qiao, L. Liu, C. Shen, and A. V. D. Hengel, "Less is more: Zero-shot learning from online textual documents with noise suppression," in *Proc. Comput. Vis. Pattern Recognit.*, 2016, pp. 2249–2257.

[29] B. Du, L. Zhang, D. Tao, and D. Zhang, "Unsupervised transfer learning for target detection from hyperspectral images," *Neurocomputing*, vol. 120, no. 10, pp. 72–82, 2013.

[30] Q. Shi, L. Zhang, and B. Du, "Domain adaptation method with low rank reconstruction and instance weighting label propagation for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 1–13, Oct. 2015.

[31] X. Li, L. Zhang, B. Du, L. Zhang, and Q. Shi, "An iterative reweighting heterogeneous transfer learning framework for supervised remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 2022–2035, May 2017.

[32] X. Zhang, "Non-negative low rank and sparse graph for semi-supervised learning," in *Proc. Comput. Vis. Pattern Recognit.*, 2012, pp. 2328–2335.

[33] V. M. Patel and H. V. Nguyen, "Latent space sparse subspace clustering," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2014, pp. 225–232.

[34] X. Wang, X. Guo, Z. Lei, C. Zhang, and S. Z. Li, "Exclusivity-consistency regularized multi-view subspace clustering," in *Proc. Comput. Vis. Pattern Recognit.*, 2017, pp. 1–9.

[35] M. Meng and J. Yu, "Zero-shot learning via robust latent representation and manifold regularization," *IEEE Trans. Image Process., submitted* for publication.

[36] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proc. Int. Conf. Neural Inf. Process. Syst.: Natural Synthetic*, 2002, pp. 585–591.

[37] T. Zhou, W. Bian, and D. Tao, "Divide-and-conquer anchoring for near-separable nonnegative matrix factorization and completion in high dimensions," in *Proc. Int. Conf. Data Mining*, 2013, pp. 917–926.

[38] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," *Adv. Neural Inf. Process. Syst.*, pp. 612–620, 2011.

[39] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.

[40] U. V. Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.

[41] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.