Research Progress of Zero-Shot Learning Beyond Computer Vision *

Weipeng Cao[1], Cong Zhou[1], Yuhao Wu[1], Zhong Ming[1], Zhiwu Xu*, and Jiyong Zhang[2]

[1] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China
[2] School of Automation, Hangzhou Dianzi University, Hangzhou, China
Email: xuzhiwu@szu.edu.cn

Abstract. Traditional machine learning techniques, including deep learning, most assume that the classes of testing samples belong to the subset of training samples. However, there are many scenarios that conflict with this assumption in the real world, that is, the classes of testing samples have never been seen in model training. To improve the generalization ability of the model in these cases, zero-shot learning (ZSL) was proposed, which can mine the mapping relationship between the features and the labels of the seen class samples and then transfer it to the prediction of unseen classes. Most of the existing ZSL algorithms or applications are concerned with computer vision problems. In fact, the above difficulties and the demand for ZSL also exist in other fields, but there is currently a lack of relevant research progress review. To make up for this gap, this paper reviews the latest research progress of ZSL beyond computer vision, introduces the general concepts of ZSL, classifies the mainstream models, and refines three issues worthy of study. This study is expected to provide ZSL-based solution guidance for researchers and engineers beyond the field of computer vision.

Keywords: Zero-shot learning· deep learning · transfer learning.

1 Introduction

In recent years, with the rapid development of machine learning techniques, especially deep learning, related algorithms have made breakthroughs in many fields, such as computer vision and natural language processing. Most of these algorithms assume that the application environment is in a closed set state, that is, the classes of the testing samples must be the classes that have been seen during the model training. However, the real world is actually an open set state, that is, sometimes the classes of testing samples are never seen by the model. We call the classes that one can see during the model training as the seen classes,

* This work was supported by National Natural Science Foundation of China (61836005) and the Opening Project of Shanghai Trusted Industrial Control Platform (TICPSH202003008-ZC) (* Corresponding author)

which corresponds to the unseen classes. Most of the traditional machine learning algorithms are devoted to the prediction of the seen classes but there is no way to predict the unseen classes. To improve the prediction performance of the model for the unseen classes, zero-shot learning (ZSL) was proposed [34], which refers to the technology that can make the model accurately predict the unseen classes [35]. According to different testing settings, existing ZSL algorithms can be divided into two categories: traditional ZSL and generalized ZSL [45]. The difference between them is that the testing samples of the former can only come from unseen classes, while the testing samples of the latter can come from both the unseen classes and the seen classes.

One of the differences between ZSL and traditional machine learning is the construction of the training data set. In ZSL, in addition to labels (i.e., the classes), it is often necessary to provide the side information corresponding to the labels. The side information is usually the semantic coding of the classes and their attributes, which can be extracted manually or by automatic techniques such as word2vec [29]. After the training data set is constructed, the ZSL algorithm learns the general knowledge in the problem domain by mining the relationship between the features of the seen class samples, class information, and the side information, and then applies it to the prediction of unseen classes. For example, for image classification problems, in the training phase, the ZSL algorithm first learns the mapping function between the visual space corresponding to the seen classes images and the semantic space corresponding to the classes; in the testing phase, given a testing sample, the ZSL model can predict the semantic feature corresponding to its class according to its visual features, then compare the semantic features corresponding to the original side information to find the closest one, and finally map back to the corresponding class to complete the prediction.

The advantages of ZSL include: (1) It can improve the generalization ability and practicality of machine learning models. For a well-trained ZSL model, it can make an accurate prediction even if it encounters unseen classes, which is critical for real-world applications; (2) it greatly eases the dependence of machine learning algorithms on labeled data. In many scenarios, the cost of collecting a large number of training samples and labeling them is very expensive, such as medical image recognition and wild animal recognition scenarios. For these cases, sometimes we can obtain some prior knowledge related to the classes in advance, which can be used to construct the side information of the classes, and then one can use ZSL to train a model with good generalization ability. ZSL has attracted extensive attention in recent years, and many related algorithms and applications have been proposed [35, 51].

However, most of the existing work is oriented to the field of computer vision. In fact, the problem to be solved by ZSL is a common problem in the real world. In other words, ZSL can be applied to solve problems other than computer vision. Actually, there have been many research achievements on this issue in recent years [30, 14, 36]. However, to the best of our knowledge, there is currently no relevant survey to introduce the research progress of ZSL beyond computer

vision. To fill this gap, this paper introduces ZSL from the following four aspects (also shown in Fig. 1):

(1) General concepts: the history of ZSL, commonly used data sets, evaluation standards, and the strategies for constructing the side information;
(2) Mainstream algorithms: categories and characteristics of mainstream ZSL models;
(3) Research progress: mainly focus on the notable work of ZSL beyond computer vision;
(4) Future trends: discussion on the future research direction of ZSL.

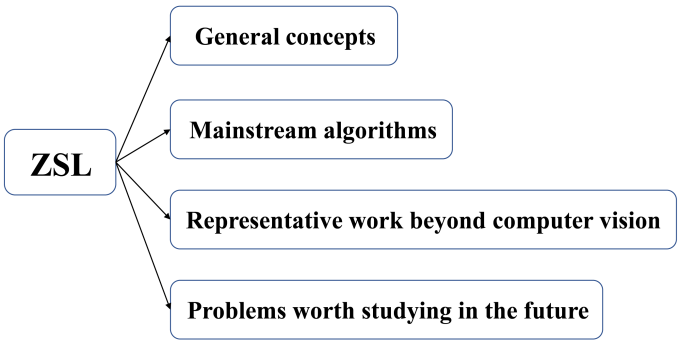

Fig. 1. The structure of this survey

Compare with other existing reviews on ZSL [35, 51], our paper has the following highlight: This is the first review of research progress beyond computer vision that provides researchers and engineers in a wider field with a basic introduction to ZSL and representative application examples, which is expected to provide them with valuable guidelines for solving a wider range of real engineering problems.

The remainder of this paper will be organized as follows. In Sec. 2, we introduce several general concepts of ZSL, such as its evaluation standards; we group some notable ZSL models into five categories in Sec. 3; Sec. 2 and Sec. 3 can help readers understand the research status of ZSL from a macro perspective. In Sec. 4, we pay attention to the representative work of ZSL beyond computer vision. In Sec. 5, we summarize three worthy research directions for researchers. We conclude this paper in Sec. 6.

2 Introduction to general concepts of ZSL

In this section, we briefly introduce the history of ZSL, the commonly used data sets, the evaluation criteria, and the constructive methods of the side information.

2.1 The proposal and development of ZSL

Before the concept of ZSL was formally proposed, some scholars had certain assumptions on the recognition of unseen classes. Representative work includes: zero-data learning of new tasks [23], ZSL with semantic output codes [34], unseen class learning by between-class attribute transfer [21], and ZSL through cross-modal transfer [40].

Specifically, zero-data learning of new tasks was proposed by Larochelle et al. in 2008 [23]. Its main goal is to apply the learned knowledge to the prediction of classes or tasks without training data, and to provide related semantic information for them, which is similar to the definition of ZSL. The authors also proposed the input space-based method and the model space-based method for dealing with this problem, which inspires the ZSL to use the side information of classes for classification.

In 2009, Palatucci et al. [34] proposed to use the semantic output coding classifier and the label base containing semantic knowledge to realize ZSL, which is the first time to propose the concept of "zero-shot learning". The model is mainly classified by the semantic codes information of class labels in the knowledge base. It compares the semantic codes information corresponding to the testing sample with the semantic codes information of the known classes in the knowledge base to distinguish the seen classes and the unseen classes.

Based on the semantic coding, Lampert et al. [21] proposed an unseen class learning method based on the between-class attribute transfer. The contributions of their work include: (1) Provided a unified training framework for most of the current ZSL methods; (2) Established a benchmark data set for ZSL (i.e., "Animals with attributes"); (3) Introduced the concept of "attribute" into ZSL. Based on this framework, two classic ZSL models are derived: the direct attribute prediction model (DAP) and the indirect attribute prediction model (IAP).

In 2013, Socher et al. [40] proposed a ZSL algorithm with a cross-modal transfer function, which transformed the ZSL into a subspace learning problem. The core idea of this method is to map the training images and their labels into the same subspace, and then use similarity measurement techniques to determine the labels of the testing samples.

Later, with the rapid development of deep learning, researchers began to use related technologies to realize the evolution from the low-level visual features to the deep-level visual features, and then better mining the mapping relationship between the visual space and the semantic space. For example, one can use the deep convolutional neural network proposed by Alex et al. in 2012 [20] and the word2vec technique proposed by Mikolov et al. in 2013 [29] to extract the deep visual features from the training data and obtain the semantic vectors corresponding to the labels, respectively, and then train the model based on the existing ZSL algorithm.

The classes of testing samples in conventional ZSL are completely different from those in the training phase, which deviates from the real-world rules because the classes of the testing samples in the real world should include both the seen classes and the unseen classes. To make the ZSL model more consistent with the

real world, generalized ZSL (GZSL) was proposed [7], which has no restrictions on the types of testing samples. Therefore, GZSL is more difficult but also more practical, which has become one of the current research hotspots.

2.2 Data sets and evaluation criteria

At present, ZSL is mainly used to solve computer vision problems and some representative data sets for this scenario include: ImageNet [10], Animals with Attributes (AwA) [7], Caltech-UCSD-Birds200-2111 (CUB) [43], Attribute Pascal and Yahoo (aPY) [12], SUN attribute [37], Oxford Flowers (FLO)[31], etc.

In the early stage, there was no unified standard to divide the seen classes and the unseen classes for a given data set, resulting in an unfair phenomenon in the performance evaluation of the ZSL algorithm. To alleviate this problem, Xian et al. proposed a standard data set segmentation method [47] in 2017, which unifies the benchmarks of model evaluation by unifying the evaluation protocol and the data segmentation. In addition, due to the uneven distribution of the classes in ZSL, the traditional mean average precision (MAP) can not reflect the performance of the ZSL algorithm well, so the class average accuracy was proposed [45] and has become one of the most commonly used evaluation indicators, which can be obtained using the following formula:

$$M = \frac{\sum_{i=1}^{K} A_u^i}{K} \tag{1}$$

where K is the number of the unseen classes and A_u^i refers to the prediction accuracy of the model on the i-th unseen class.

Moreover, to better evaluate the performance of the GZSL algorithm, the harmonic mean $\mathcal{H}$ was proposed [47] and has become one of the most commonly used evaluation indicators, which can be obtained as follows:

$$\mathcal{H} = \frac{2 \times A_s \times A_u}{A_s + A_u} \tag{2}$$

where A_s and A_u are the top-1 accuracy of the model on the seen classes and the unseen classes, respectively.

2.3 Side information

Side information is used to describe the auxiliary information of the classes such as their attributes. It serves as a bridge between the seen classes and the unseen classes, making it possible to use the seen samples to train a ZSL model that can predict the unseen classes. Therefore, side information is an important part of ZSL, which can usually be obtained through two methods: human annotation and text-based learning [35, 51].

Methods based on human annotation can be further divided into the attribute-based method and the non-attribute-based method. As a kind of prior knowledge,

attributes can semantically represent specific classes and reflect their characteristics, so that they can be used to distinguish different classes in a data set. Commonly used attribute representation can be divided into binary attributes and continuous attributes. Binary attributes are used to describe whether a specific class or object has a certain attribute, if it is, the value of this dimension is 1, otherwise it is 0. Continuous attributes are generally used to describe the possibility of a specific class or object having a specific attribute, which is usually expressed in the form of real values. The non-attribute-based method directly uses the names of the classes as their semantic information description to construct the semantic vectors. Methods based on human annotation need the help of human's prior knowledge, so it inevitably has certain subjectivity. Moreover, when there are many classes, this method is time-consuming and expensive. The advantage of this method is that it is helpful to get high accuracy of the ZSL model, and to some extent, it can improve the interpretability of the results.

Text-based learning methods mainly use machine learning algorithms to obtain mapping models and then map classes or their descriptions to corresponding vectors. According to different auxiliary information, this kind of method can be divided into two categories: label embedding and text embedding. Label embedding mainly uses natural language processing models, such as word2vec, to represent the class labels in vectorization. The similarity between word vectors can be used as a reference for classification. The text embedding method needs to obtain the description text of the classes, and convert the text description into the corresponding semantic vector through text encoding models. The text-based methods can effectively reduce the labor cost, but the disadvantage is that the data sources often have noise and the results are less interpretable.

3 Categories and characteristics of mainstream models

Inspired by the classification method of [45], here we divide the existing notable ZSL models into five categories: intermediate attribute classifiers models, compatibility models, hybrid models, transductive models, and generative models. The details of these five categories and the corresponding notable algorithms are shown in Table 1.

3.1 Intermediate attribute classifiers models

In this learning paradigm, attributes are the key information that the ZSL model uses to make decisions. Specifically, given a testing sample, the ZSL model first predicts the attribute of its class and then selects the most probable class according to the similarity of the attribute to the attributes of the known classes. The consistency model and the hybrid model are also derived from the intermediate attribute classifiers models.

At present, the existing intermediate attribute classifiers models can be divided into two categories: the direct attribute prediction model (DAP) and the indirect attribute prediction model (IAP) [22]. In the training phase, DAP trains

Table 1. The details of the five categories of ZSL and the corresponding notable algorithms

Categories	**Notable algorithms**
Intermediate attribute classifiers models	DAP [27], IAP [22], etc.
Compatibility models	DEVISE [15], ALE [1], SJE [2], ES-ZSL [39], SAE [19], LATEM [44], CMT [40], etc.
Hybrid models	SSE [52], CONSE [32], SYNC [6], GFZSL [42], etc.
Transductive models	GFZSL-tran [42], DSRL [50], etc.
Generative models	f-xGAN [46], cycle-CLSWGAN [13], AFC-GAN [25], etc.

the attribute classifier. In the testing phase, one can directly obtain the attribute feature estimate by inputting the attributes of the testing sample into the model even if the testing class is an unseen class. The difference between IAP and DAP is that the attribute classifier of IAP cannot directly obtain the attribute feature estimate. IAP needs to input the class label of the sample and its attribute indication vector to indirectly obtain the attribute feature estimate.

3.2 Compatibility models

The compatibility models are models that map the input and output to a subspace and then judge the compatibility of the input and output mapping vectors in the subspace and determine the class label. The compatibility models can be divided into linear compatibility models and non-linear compatibility models according to whether the compatibility function is linear. Some notable linear compatibility models include: deep visual-semantic embedding model (DEVISE) [15], attribute label embedding model (ALE) [1], structured joint embedding (SJE) [2], embarrassingly simple zero-shot learning (ESZSL) [39], semantic auto-encoder embedding (SAE) [19], etc. Some representative non-linear compatibility models include: latent embedding models (LATEM) [44], cross-mode migration model (CMT) [40], etc.

Compatibility models need to train the class label embedding function so that the class label can be accurately embedded in the feature space. However, the compatibility models rely heavily on the quality of auxiliary information. The learning ability of the linear compatibility models is often limited by the linear function, so its expression ability is not as good as the nonlinear compatibility models.

3.3 Hybrid models

The hybrid models use a hybrid combination of feature subspace mappings corresponding to the training class labels to represent the mapping of the testing samples in the feature subspace, and then obtain the class label estimation of the

testing sample based on the similarity between the input sample mapping and the testing class label mapping. Some notable hybrid models include: semantic similarity embedding model (SSE) [52], convex combination semantic embedding model (CONSE) [32], synthesized classifiers (SYNC) [6], generative framework for zero-shot learning (GFZSL) [42], etc.

The training mechanism of the hybrid models is similar to that of the IAP, so they also rely heavily on the similarity between the training classes and the testing classes, and the robustness of their models is relatively weak.

3.4 Transductive models

Transductive models [41] use the class labels of the training classes and the side information of the testing classes to determine the class labels of the testing samples, and then add the testing samples with their predicted class labels to the original training data set. Then, continue to learn the new decision rules based on the augmented data set, and use the updated model to continue the above steps until all testing samples are labeled. Representative transductive models include GFZSL-tran [42], discriminative semantic representation learning (DSRL) [50], etc.

Transductive models belong to an online learning paradigm that can continue to update themselves during the testing phase. The disadvantage of this kind of method is that their training process needs a lot of calculation and the final model performance dependents on the initial accuracy of the model.

3.5 Generative models

Recently, it is popular to combine generative adversarial networks (GAN) [18] with ZSL to produce generative ZSL models. Using the side information of the classes as the constraint of the GAN model can enable its generator to generate features related to the specific class, which in turn can make the model better distinguish different classes.

In particular, if the corresponding pseudo samples or features can be generated according to the side information of the unseen classes, the ZSL task can be converted into traditional supervised learning. Some notable work includes: f-xGAN [46], cycle-CLSWGAN [13], alleviating feature confusion GAN (AFC-GAN) [25], etc.

4 Application progress of ZSL beyond computer vision

ZSL has been widely used in the field of computer vision and some researchers have reviewed these works well [35, 51]. Different from the existing surveys, here we focus on the research progress of ZSL in natural language processing and other fields.

4.1 Application of ZSL in natural language processing

Nakashole et al. [30] combined ZSL with the bilingual dictionary induction to realize that one can translate an uncommon language (e.g., Portuguese) into a common language (e.g., English) through a third relatively relevant language (e.g., Spanish) when only a small seed dictionary is used. The authors claimed that using ZSL to train the translation model can achieve high accuracy even with a small amount of labeled data.

In [14], Ferreira et al. proposed a complete semantic analyzer based on the word embedding and ZSL techniques. This semantic analyzer does not require annotated contextual data, ontology description of the target domain, and general word embedding features, which reduces the cost of manual annotation and can obtain performance comparable to the use of manual annotation data.

Pasupat et al. [36] used the ZSL technique to extract entities of specified categories in web pages. Most of the traditional methods are effective for ordinary categories, but they are relatively weak when facing categories with few samples. ZSL does not rely on multiple web pages to get entities, it only needs a single web page to get the target entities. Moreover, it can extract the target entities from the semi-structured data on the web page without complete category information.

In [28], Ma et al. used a ZSL framework to solve the problem that current named entity recognition methods cannot detect unseen entities. They proposed a label embedding method that combines prototype information and hierarchical information to learn the pre-trained label embedding. In this way, the above problems can be alleviated to some extent.

Funaki et al. [17] proposed an image-based cross-language document retrieval method, which can take images in two languages as the target, and deduce the common semantic subspace connecting two languages through generalized canonical correlation analysis, to realize document retrieval between different languages. This method can reduce the cost of manually creating a corpus when there is no or only a small number of parallel corpora.

4.2 Application of ZSL in other fields

In addition to applications in computer vision and natural language processing, ZSL has also been used in other fields. For example, for the human activity recognition problem, Zheng et al. [53] proposed that existing experience can be used to identify unseen human activities through knowledge transfer methods, and then Cheng et al. [9] realized the recognition of unseen activity categories based on the semantic description by using ZSL.

In the field of knowledge representation, Xie et al. [48] proposed the DKRL model to deal with the task of entity classification where at least one entity in the triple is not in the knowledge graph.

Robyns et al. [38] realized the identification of unseen physical layer devices by using ZSL. ZSL can also complete tasks such as generating Emoji expressions for unseen images [11], neural decoding of fMRI images [3], and identifying unseen molecular compounds [23].

5 Future trends

Here we propose three open problems worthy of study..

5.1 Smart side information construction method

As mentioned in Sec. 2.3, the side information generated by manual labeling has the advantages of high accuracy and strong interpretability, but its disadvantages are high cost and low efficiency. The automatic text-based learning method can overcome these shortcomings, but the quality of the side information obtained by this method is difficult to guarantee. Combining these two methods to design a smart side information construction method will be a problem worthy of study. It is expected that the new method can obtain abundant side information quickly and accurately.

5.2 Generalized zero-shot learning

As mentioned in 2.1, GZSL is more in line with the rules of the real world, that is, the testing samples can come from both the seen and unseen classes. Since the unseen classes samples have never been seen during model training, it is very difficult for the model to accurately predict them. Moreover, in this case, the three challenging issues in the ZSL field (i.e., domain shift problem [16], hubness problem [24], and the semantic gap problem [26]) will become more difficult [8]. How to improve the generalization ability of the ZSL model under the evaluation setting of GZSL is one of the current research hotspots. Although some researchers have put forward some enlightening algorithms [8, 32, 7, 49], there is still a long way to go before GZSL can be truly applied. How to improve the generalization ability of the GZSL model is one of the most worthy research issues.

5.3 Combination with other technologies

Most fields are facing the problem of open set learning, how to combine the idea of ZSL with the existing algorithms in these fields to improve their generalization ability is a direction worth exploring. For example, some researchers have combined ZSL with reinforcement learning to optimize the strategies of agents to deal with unknown environments [33]. Moreover, one can also learn from the advantages of other technologies to improve the performance of the ZSL algorithm. For example, neural networks with random weights (NNRW [5, 4]) have extremely fast training speeds, how to combine them with the existing ZSL algorithms to improve the training efficiency of the latter is an interesting research direction.

6 Conclusions

In this paper, we first introduce the research motivation, development history, and fundamental concepts of ZSL, and then classify the current mainstream ZSL models into five categories, which enable readers to have a macro understanding of the field of ZSL. Furthermore, we focus on the representative work of ZSL beyond computer vision, which is the biggest difference from other existing surveys. Moreover, we have refined three valuable research issues to provide direction for researchers. This paper is expected to provide guidance on open set learning for researchers and engineers in a wider range of fields.

In the future, we will give more details for the difficult issues mentioned in this paper, and add more representative work of ZSL beyond computer vision.

References

1. Akata, Z., Perronnin, F., Harchaoui, Z., Schmid, C.: Label-embedding for image classification. IEEE Transactions on Pattern Analysis and Machine Intelligence **38**(7), 1425–1438 (2016)
2. Akata, Z., Reed, S., Walter, D., Lee, H., Schiele, B.: Evaluation of output embeddings for fine-grained image classification. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2927–2936 (2015)
3. Caceres, C.A., Roos, M.J., Rupp, K.M., Milsap, G., Crone, N.E., Wolmetz, M.E., Ratto, C.R.: Feature selection methods for zero-shot learning of neural activity. Frontiers in neuroinformatics **11**, 41 (2017)
4. Cao, W., Hu, L., Gao, J., Wang, X., Ming, Z.: A study on the relationship between the rank of input data and the performance of random weight neural network. Neural Computing and Applications pp. 1–12 (2020)
5. Cao, W., Wang, X., Ming, Z., Gao, J.: A review on neural networks with random weights. Neurocomputing **275**, 278–287 (2018)
6. Changpinyo, S., Chao, W.L., Gong, B., Sha, F.: Synthesized classifiers for zero-shot learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5327–5336 (2016)
7. Chao, W.L., Changpinyo, S., Gong, B., Sha, F.: An empirical study and analysis of generalized zero-shot learning for object recognition in the wild. In: European Conference on Computer Vision. pp. 52–68 (2016)
8. Chao, W.L., Changpinyo, S., Gong, B., Sha, F.: An empirical study and analysis of generalized zero-shot learning for object recognition in the wild. In: European Conference on Computer Vision. pp. 52–68. Springer (2016)
9. Cheng, H.T., Sun, F.T., Griss, M., Davis, P., Li, J., You, D.: Nuactiv: Recognizing unseen new activities using semantic attribute-based learning. In: Proceeding of the 11th annual international conference on Mobile systems, applications, and services. pp. 361–374 (2013)
10. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
11. Dinu, G., Lazaridou, A., Baroni, M.: Improving zero-shot learning by mitigating the hubness problem. In: ICLR (Workshop) (2014)

12. Farhadi, A., Endres, I., Hoiem, D., Forsyth, D.: Describing objects by their attributes. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1778–1785 (2009)
13. Felix, R., Kumar, V.B., Reid, I., Carneiro, G.: Multi-modal cycle-consistent generalized zero-shot learning. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 21–37 (2018)
14. Ferreira, E., Jabaian, B., Lefèvre, F.: Zero-shot semantic parser for spoken language understanding. In: Sixteenth Annual Conference of the International Speech Communication Association (2015)
15. Frome, A., Corrado, G.S., Shlens, J., Bengio, S., Dean, J., Ranzato, M., Mikolov, T.: Devise: A deep visual-semantic embedding model. In: Advances in Neural Information Processing Systems 26. pp. 2121–2129 (2013)
16. Fu, Y., Hospedales, T.M., Xiang, T., Fu, Z., Gong, S.: Transductive multi-view embedding for zero-shot recognition and annotation. In: European Conference on Computer Vision. pp. 584–599. Springer (2014)
17. Funaki, R., Nakayama, H.: Image-mediated learning for zero-shot cross-lingual document retrieval. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. pp. 585–590 (2015)
18. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems. pp. 2672–2680 (2014)
19. Kodirov, E., Xiang, T., Gong, S.: Semantic autoencoder for zero-shot learning. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4447–4456 (2017)
20. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012)
21. Lampert, C.H., Nickisch, H., Harmeling, S.: Learning to detect unseen object classes by between-class attribute transfer. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 951–958 (2009)
22. Lampert, C.H., Nickisch, H., Harmeling, S.: Attribute-based classification for zero-shot visual object categorization. IEEE Transactions on Pattern Analysis and Machine Intelligence **36**(3), 453–465 (2014)
23. Larochelle, H., Erhan, D., Bengio, Y.: Zero-data learning of new tasks. In: AAAI'08 Proceedings of the 23rd national conference on Artificial intelligence - Volume 2. pp. 646–651 (2008)
24. Lazaridou, A., Dinu, G., Baroni, M.: Hubness and pollution: Delving into cross-space mapping for zero-shot learning. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). vol. 1, pp. 270–280 (2015)
25. Li, J., Jing, M., Lu, K., Zhu, L., Yang, Y., Huang, Z.: Alleviating feature confusion for generative zero-shot learning. In: Proceedings of the 27th ACM International Conference on Multimedia. pp. 1587–1595 (2019)
26. Li, Y., Wang, D., Hu, H., Lin, Y., Zhuang, Y.: Zero-shot recognition using dual visual-semantic mapping paths. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5207–5215 (2017)
27. Liang, K., Chang, H., Shan, S., Chen, X.: A unified multiplicative framework for attribute learning. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2506–2514 (2015)

28. Ma, Y., Cambria, E., Gao, S.: Label embedding for zero-shot fine-grained named entity typing. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. pp. 171–180 (2016)
29. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)
30. Nakashole, N., Flauger, R.: Knowledge distillation for bilingual dictionary induction. In: Proceedings of the 2017 conference on empirical methods in natural language processing. pp. 2497–2506 (2017)
31. Nilsback, M.E., Zisserman, A.: Automated flower classification over a large number of classes. In: 2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing. pp. 722–729. IEEE (2008)
32. Norouzi, M., Mikolov, T., Bengio, S., Singer, Y., Shlens, J., Frome, A., Corrado, G.S., Dean, J.: Zero-shot learning by convex combination of semantic embeddings. arXiv preprint arXiv:1312.5650 (2013)
33. Oh, J., Singh, S., Lee, H., Kohli, P.: Zero-shot task generalization with multi-task deep reinforcement learning. In: Proceedings of the 34th International Conference on Machine Learning-Volume 70. pp. 2661–2670. JMLR. org (2017)
34. Palatucci, M., Pomerleau, D., Hinton, G.E., Mitchell, T.M.: Zero-shot learning with semantic output codes. In: Advances in Neural Information Processing Systems 22. vol. 22, pp. 1410–1418 (2009)
35. Pang, Y., Wang, H., Yu, Y., Ji, Z.: A decadal survey of zero-shot image classification. SCIENTIA SINICA Informationis **49**(10), 1299–1320 (2019)
36. Pasupat, P., Liang, P.: Zero-shot entity extraction from web pages. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 391–401 (2014)
37. Patterson, G., Hays, J.: Sun attribute database: Discovering, annotating, and recognizing scene attributes. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 2751–2758 (2012)
38. Robyns, P., Marin, E., Lamotte, W., Quax, P., Singelée, D., Preneel, B.: Physical-layer fingerprinting of lora devices using supervised and zero-shot learning. In: Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks. pp. 58–63 (2017)
39. Romera-Paredes, B., Torr, P.: An embarrassingly simple approach to zero-shot learning. In: Proceedings of The 32nd International Conference on Machine Learning. pp. 2152–2161 (2015)
40. Socher, R., Ganjoo, M., Sridhar, H., Bastani, O., Manning, C.D., Ng, A.Y.: Zero-shot learning through cross-modal transfer. In: ICLR (Workshop) (2013)
41. Song, J., Shen, C., Yang, Y., Liu, Y., Song, M.: Transductive unbiased embedding for zero-shot learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1024–1033 (2018)
42. Verma, V.K., Rai, P.: A simple exponential family framework for zero-shot learning. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 792–808. Springer (2017)
43. Wah, C., Branson, S., Welinder, P., Perona, P., Belongie, S.: The caltech-ucsd birds-200-2011 dataset (2011)
44. Xian, Y., Akata, Z., Sharma, G., Nguyen, Q., Hein, M., Schiele, B.: Latent embeddings for zero-shot classification. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 69–77 (2016)
45. Xian, Y., Lampert, C.H., Schiele, B., Akata, Z.: Zero-shot learning-a comprehensive evaluation of the good, the bad and the ugly. IEEE transactions on pattern analysis and machine intelligence **41**(9), 2251–2265 (2018)

46. Xian, Y., Lorenz, T., Schiele, B., Akata, Z.: Feature generating networks for zero-shot learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5542–5551 (2018)
47. Xian, Y., Schiele, B., Akata, Z.: Zero-shot learning — the good, the bad and the ugly. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3077–3086 (2017)
48. Xie, R., Liu, Z., Jia, J., Luan, H., Sun, M.: Representation learning of knowledge graphs with entity descriptions. In: Thirtieth AAAI Conference on Artificial Intelligence (2016)
49. Xie, Z., Cao, W., Wang, X., Ming, Z., Zhang, J., Zhang, J.: A biologically inspired feature enhancement framework for zero-shot learning. arXiv preprint arXiv:2005.08704 (2020)
50. Ye, M., Guo, Y.: Zero-shot classification with discriminative semantic representation learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7140–7148 (2017)
51. Zhang, L.N., Zuo, X., Liu, J.W.: Research and development on zero-shot learning. Acta Automatica Sinica **46(1)**(46), 1 (2020)
52. Zhang, Z., Saligrama, V.: Zero-shot learning via semantic similarity embedding. In: Proceedings of the IEEE international conference on computer vision. pp. 4166–4174 (2015)
53. Zheng, V.W., Hu, D.H., Yang, Q.: Cross-domain activity recognition. In: Proceedings of the 11th international conference on Ubiquitous computing. pp. 61–70 (2009)