Research Article/Araştırma Makalesi

GAZİ
JOURNAL OF ENGINEERING SCIENCES

Ontology-Based Generalized Zero-Shot Learning with Generative Networks

Emre Akdemir[*a], Necattin Barışçı[b]

Submitted: 15.12.2023 Revised: 17.04.2024 Accepted: 25.04.2024 doi:10.30855/gmbd.0705N15

Keywords: Generalized zero-shot learning, Natural language processing, Ontology, Variational autoencoder, Generative adversarial network

[a,*] Gazi University, Graduate School of Informatics, Dept. of Computer Science 06500 - Ankara, Türkiye
Orcid: 0000-0003-2507-9264
e mail: emre.akdemir@gazi.edu.tr

[b] Gazi University, Technology Faculty, Dept. of Computer Engineering 06560 - Ankara, Türkiye
Orcid: 0000-0002-8762-5091

[*]Corresponding author: emre.akdemir@gazi.edu.tr

ABSTRACT

Zero-Shot Learning (ZSL) aims to classify images of new categories in the testing phase without labeled images during training, using examples from categories with labeled images and some auxiliary information. The auxiliary information includes semantic attributes, textual descriptions, word embeddings, etc., for both labeled and unlabeled classes, utilizing Natural Language Processing (NLP) approaches. The word embeddings created are limited by the semantic attributes and textual descriptions where the semantics of categories are insufficient. In this paper, introduces a study for Generalized Zero-Shot Learning (GZSL), a type of ZSL, by integrating the rich semantics offered by ontology. Semantic attributes used for semantic representation are supported by ontology. Variational Autoencoder (VAE) and Generative Adversarial Network (GAN) network architectures are used together to synthesize visual features. Our work was evaluated on the AWA2 dataset, and improvement in GZSL performance was achieved.

Üretici Ağlar ile Ontoloji Tabanlı Genelleştirilmiş Sıfır-Atışlı Öğrenme

ÖZ

Sıfır-Atışlı Öğrenme (Zero-Shot Learning - ZSL), eğitim sırasında, etiketli görüntülerin bulunduğu kategorilere ait örneklerden ve bazı yardımcı bilgilerden yararlanarak test aşamasında etiketli görüntüleri bulunmayan yeni kategorilere ait örnekleri sınıflandırmayı amaçlamaktadır. Buradaki yardımcı bilgiler hem etiketli hem de etiketsiz sınıflar için semantik öznitelikler, metinsel açıklamalar, sözcük gömme gibi doğal dil işleme yaklaşımlarıdır. Oluşturulan sözcük gömmeleri, kategorilerin anlambiliminin yetersiz olduğu semantik öznitelikler ve metinsel açıklamalar ile kısıtlıdır. Bu yazıda ontolojinin sunduğu zengin semantiği üretici ağlara entegre ederek ZSL'nin bir türü olan Genelleştirilmiş Sıfır-Atışlı Öğrenme (Generalized Zero-Shot Learning - GZSL) görevi için bir çalışma tanıtılmıştır. Semantik temsil için kullanılan semantik öznitelikleri ontoloji ile desteklenmiştir. Görsel özellikleri sentezlemek için VAE ve GAN ağlarını birlikte kullanılmıştır. Çalışmamızı AWA2 veri seti üzerinde değerlendirilmiştir ve GZSL performansında iyileştirme sağlanmıştır.

Anahtar Kelimeler: Genelleştirilmiş sıfır-atışlı öğrenme, Doğal dil işleme, Ontoloji, Varyasyonel otomatik kodlayıcı, Çekişmeli üretici ağ

To cite this article: E. Akdemir and N. Barisci, "Ontology-Based Generalized Zero-Shot Learning with Generative Networks," *Gazi Journal of Engineering Sciences*, vol.10, no.1, pp. 183-192, 2024. doi:10.30855/gmbd.0705N15

1. Introduction

Deep learning has made significant advances in image processing, but it relies on collecting sufficient labeled data for recognizing each category. In some cases, it may not be feasible to gather labeled training data for every category [1]. The absence of labeled data makes the task considerably challenging for machine learning and deep learning technologies [2] . In recent years, Zero-Shot Learning (ZSL) has gained attention for its successful results in classifying unlabeled images [3]. ZSL consists of seen classes (with labeled image examples), unseen classes (without labeled image examples), and semantic representations. ZSL aims to recognize examples from unseen classes by transferring knowledge from examples in seen classes and semantic representations[1]. During the testing phase of the ZSL task, only the performance of unseen classes is evaluated. For a more realistic evaluation, Generalized Zero-Shot Learning (GZSL) studies have started to replace ZSL studies. In the testing phase of the GZSL task, the performance of both seen and unseen classes is assessed.

Due to the absence of labeled examples for new classes, GZSL encounters a data imbalance issue between seen and unseen classes. Generative approaches, such as Generative Adversarial Networks (GAN) and Variational Autoencoders (VAE), are highly effective in addressing this problem. Using generative networks, synthetic visual features are produced for unseen classes by leveraging both visual and semantic features. However, textual features are not generated for semantic relationships and contextual information of unseen classes [4]. Therefore, there is a need for structured data, such as semantic web, to address this issue [5].

The use of various Natural Language Processing (NLP) methods in deep learning tasks is quite common [6]. Extracting meaning from textual data such as web pages and documents using deep learning methods is a challenging task. To address these challenging tasks, there is a need for rich models consisting of semantic concepts offered by NLP tasks. Successful results have been achieved in tasks such as extracting meaning from texts and establishing relationships between images and text using NLP-based deep learning methods [7-9]. Most words have multiple meanings. Relationships such as synonyms, antonyms, similar meanings, as well as positive or negative connotations can exist for a word. In short, there are relationships and representations between words. The semantic concept mentioned in NLP refers to how the meaning of a word is represented and the information obtained from its relationship with other words[10].

Semantic Web emerges from the relationships established among semantic information. The goal of the Semantic Web is to enable machines to make inferences. To achieve this, a structure is formed through the collaboration of humans and computers. This structure is obtained by comprehending and linking data on the web. Ontology is often used to create a robust structure with Semantic Web. Ontologies provide a strong structure where machines can understand relationships between concepts. An ontology can be specific to a domain or an advanced ontology resulting from the integration of ontologies from different domains. Commonly, Resource Description Framework (RDF) and Web Ontology Language (OWL) are used to describe the Semantic Web[11-13].

Fig. 1 provides an example representation for the seen class, unseen class, and ontological schema. The semantic attributes used for images contain values for specific features in a given category. The ontological schema, in addition to semantic attributes, presents a relational structure by utilizing class hierarchy and descriptions. To create the ontological schema, the class hierarchy offered by RDF, such as "subClassOf," is initially employed. Subsequently, class concepts are associated with attribute concepts like "hasTexture." As depicted in Fig. 1, class concepts are represented by ovals, and attribute concepts are represented by rectangles. Additionally, each class includes text-based descriptions. The created ontological schema is an RDF triple. For example, the triple "killer_whale," "hasTexture," "spots" expresses that the killer whale has a spotted texture. The ontological schema enhances the information transfer between seen and unseen classes more effectively.

Building on this, in our study, we aimed to enhance the performance of GZSL by enriching semantic representations. For this purpose, we augmented the semantic attributes in the datasets with an ontology created using RDF. Subsequently, we designed a VAE-GAN architecture for the learning model. We evaluated our proposed GZSL model on the AWA2 dataset. In the conducted experiments, our proposed method resulted in a 1.19% improvement in GZSL performance. To the best of our knowledge, our GZSL study is the first to incorporate RDF and semantic attributes together into a VAE-GAN-based model.

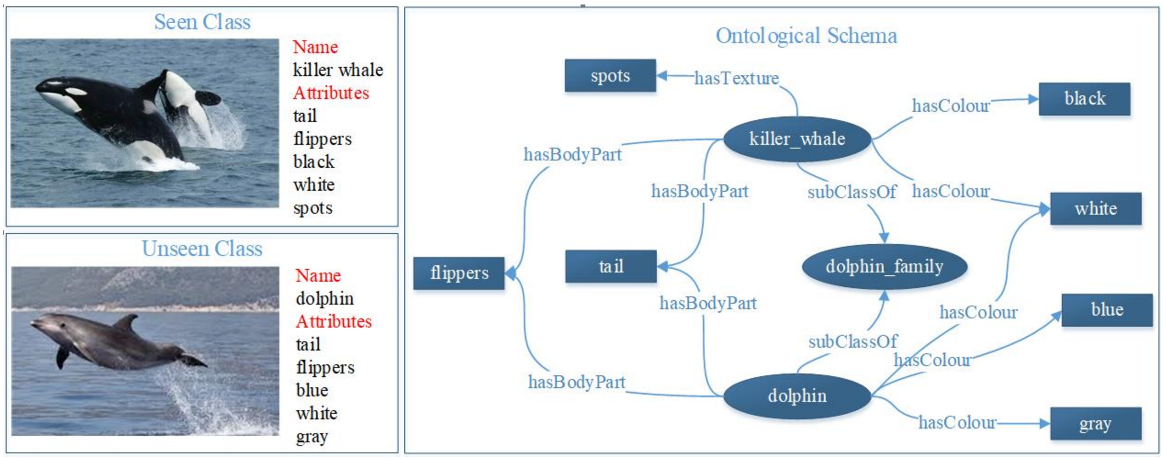

Figure 1. Semantic attribute and ontological schema representations

The rest of the article is organized as follows: Section 2 provides a summary of related works in the literature. Section 3 introduces the proposed model in detail. Section 4 presents the experimental results of the study. Finally, Section 5 concludes with conclusions and discussion.

2. Related Work

GZSL is the recognition of an object without any instance images of that object during the training phase, utilizing textual and image data from other examples. GZSL is categorized into embedding-based methods and generative-based methods. Embedding-based methods find a projection function from the visual space to the semantic space during the training phase of the GZSL method, using information from seen categories. During the testing phase, unseen image feature data is used as input to the trained model, and a semantic embedding is obtained. Classification is then performed based on the category corresponding to the nearest semantic embedding. [14-15]. Generative-based methods enable unseen classes to behave like seen classes by synthesizing synthetic features for the unseen classes. In studies using generative-based methods, GAN is first used and then VAE is employed to learn the synthesis of visual features for unseen classes from semantic attributes. In subsequent studies, a VAE-GAN model, combining the strengths of both VAE and GAN, was developed. [16]. Wu et al. utilized Stacked Autoencoder (StAE) to learn the relationship between extracted features and semantic representations [17]. Gao et al. proposed a shared generative model known as Zero-VAE-GAN to generate high-quality visual features for unseen classes. In the model, a conditionally guided VAE based on semantic features is combined with a GAN conditioned on both categories and features. Additionally, a classification network and perceptual reconstruction loss are included to generate high-quality features [18]. Narayan et al. worked on an approach for both image and video datasets by adding a feedback module to a VAE-GAN-based model [19]. Bao et al. colleagues utilized the VAE-GAN model to create realistic and diverse new examples from natural images of faces, birds, and flowers within a class [20]. Han et al. suggested a comparative embedding. In the proposed method, they used a parallel inference network and a feature generation framework to project visual features into a semantic descriptive space [21].

Semantic web technologies enhance the restructuring of information and facilitate machine readability. Semantic web technologies have been repeatedly used in machine learning and deep learning tasks. To enable a structure like ontology to be utilized by a deep neural network, there is a need for a framework that can produce multiple hierarchical outputs representing the data. Each output of the neural network corresponds to a concept in the ontology. Thus, a taxonomic assumption relationship architecture is defined for all concepts for which inferences can be made based on these outputs [22,23]. In Semantic Web-based applications, formal representation languages such as RDF and OWL, as well as structures like ontologies and knowledge graphs, are commonly used. Geng et al. proposed a new generative zero-shot learning method called KG-GAN by incorporating rich semantics from a knowledge graph into GANs. They developed graph neural networks to investigate the effects of class semantics on feature transfer in zero-shot learning and to learn semantic class nodes. They utilized the original taxonomy structure of WordNet along with the class and subclass relationships of ImageNet to create the knowledge graph [24]. Geng et al. proposed an ontology-based GAN to create more distinctive example features for ZSL. Their approach, named Onto-ZSL, demonstrated the effectiveness of ZSL on image classification and knowledge graph completion tasks [25]. In addition to the image classification and knowledge graph completion tasks introduced in their previous work, Geng et al. presented another study involving ZSL with a relation extraction task [26].

Upon examining previous studies, it is evident that the use of VAE-GAN-based generative approaches has led to highly successful results in GZSL tasks. In GZSL classification, the synthesized image features for unseen classes are crucial. At the same time, semantic relationships play a significant role. Therefore, we propose a GZSL framework that enriches semantic relationships with ontology, incorporating generative approaches to enhance GZSL classification.

3. Method

In this section, we first address the problem, and then we introduce the architecture of our approach in three sub-sections with detailed explanations.

3.1. Problem definition

GZSL aims to classify examples from both seen class S and unseen class U, using examples from the training dataset, which contains seen classes, and examples from the test dataset, which contains both seen and unseen classes. During the training phase of GZSL, both the seen class S and the semantic features A are trained, and during the testing phase, it learns examples from both the seen class S and the unseen class U. Here, the classes S and U are disjoint, meaning $S \cap U = \emptyset$. The semantic features A are represented as A^S for the semantic features of seen classes and A^U for the semantic features of unseen classes. Thus, the semantic features A are expressed as $A = \{a_{(k)}\}_{k=1}^{S+U}$. The dataset for the seen class S containing examples X^S corresponding to labels Y^S is denoted as $S = \{X^S, Y^S\}$. Similarly, the dataset for the unseen class U, containing examples X^U corresponding to labels Y^U, is denoted $U = \{X^U, Y^U\}$. While ZSL learns the classifier $f_{zsl}: X \to Y^U$I GZSL learns the classifier $f_{gzsl}: X \to Y^S \cup Y^U$.

3.2. Visual feature extraction

In this study, visual features obtained from a pre-trained Convolutional Neural Network (CNN) with the ResNet-101 architecture from ImageNet [27] were used. The ResNet-101 architecture is a CNN with a depth of 101 layers. These networks have been trained on over a million images and can classify images into 1000 object categories. Using a pre-trained network with transfer learning is often faster and easier than training a network from scratch. With this architecture, rich feature representations for a variety of images are learned, and 2048-dimensional visual features of images are extracted through average pooling. CNN has a powerful capability for extracting visual features. However, unlike traditional image recognition, GZSL involves unlabeled images, which leads to a decrease in the discriminative power of visual features extracted by CNN [28]. Therefore, for GZSL, it is not sufficient to rely solely on visual feature extraction, and semantic feature extraction is also utilized.

3.3. Semantic feature extraction

3.3.1. Semantic embedding

Semantic information is represented with semantic vector embeddings. A semantic vector is a representation of a word in a multi-dimensional semantic space. Semantic embeddings provide an effective way to represent the meaning of words and analyze the semantic aspects of language. The use of language models such as word2vec, GloVE, and BERT is common for representing the meaning of words. In recent years, the positive impact of word embeddings on the performance of many applications has been observed [10,29,30]. In traditional word embedding methods, the lack of word order information leads to a lack of semantic information for words. Therefore, there is a problem of semantic compositionality in word embeddings. To address this issue, a Recurrent Neural Network (RNN)-based language model that represents the meaning of words has been developed, resulting in more meaningful word embeddings [31,32].

In traditional GZSL studies, semantic attributes are used to create semantic embeddings. In our study, in addition to semantic attributes, textual descriptions and ontology are also utilized to create semantic embeddings.

3.3.2. Ontology

In this study, RDF (Resource Description Framework) was used to create an ontology. RDF represents web resources using triple statements, consisting of subject, predicate, and object. Subjects and objects represent entities, while predicates represent the relationships between these entities. When given the statement "Ankara is the capital of Turkey," RDF should be able to infer "Ankara" and "Turkey" as entities and the relationship "is the capital of" between them [33]. As seen, manually coding information with human input is crucial for developing semantic web-based applications [34]. Ontologies are hierarchical representations of elements in a domain. These hierarchical representations are used to define categories at each level [35]. In our study, an ontology schema containing class hierarchy, semantic attributes, and textual descriptions, created using RDF triples, was employed [25]. The statistics of the created ontological schemas are shown in Table 1. In Table 1, concepts refer to entities such as class names and class attributes. Properties represent relationships established between concepts, such as color, texture, subclass, etc. RDF triples consist of three elements, two concepts, and a property.

Table 1. Ontological schema information for the AWA2 dataset

Dataset	RDF Triples	Concepts	Properties
AWA2	1256	180	12

In GZSL applications, the input dimensions of generative networks are determined by the dimensions of visual features and semantic attributes. The visual and semantic feature dimensions of the AWA2 dataset used in our study are shown in Table 2. The visual features of the images are 2048-dimensional. For semantic embeddings, 85-dimensional class-level attributes, 100-dimensional structure-based representation, and 100-dimensional text-based representation are utilized.

Table 1. Feature dimensions for the AWA2 dataset

Dataset	Visual Features	Semantic Attributes	RDF	Textual Descriptions
AWA2	2048	85	100	100

3.4. Generative-based networks

The examples generated by VAE are often blurry. Similarly, the examples generated by GAN are not natural. However, the combination of VAE and GAN produces much more realistic, natural images [20]. In this study, the VAE-GAN-based model includes Encoder, Generator, Discriminator, and Decoder networks. All networks in the model have 4096 hidden units and are two-layer fully connected networks. The proposed architecture is shown in Fig. 2. In the workflow, first, the real features of seen classes x and semantic embedding a are applied to the input of the Encoder, and the parameters of a noise distribution z are extracted. Then, the extracted z and semantic embedding a are applied to the input of the Generator, and synthetic features $\hat{x}$ are generated. Synthetic features $\hat{x}$ and original features x are compared using Binary Cross-Entropy loss (L_{BCE}). Up to this point, it is the standard VAE approach. Encoder and Generator, which constitute the VAE, are trained using the L_{BCE} and the Kullback-Leibler divergence (L_{KL}). The synthetic feature $\hat{x}$ or the original feature x, along with semantic embedding a, is applied to the input of the Discriminator, and the Discriminator learns to distinguish synthetic and original features. Here, Generator and Discriminator follow the standard GAN approach. Generator and Discriminator are trained using the Wasserstein GAN loss (L_W). To improve classification performance, the synthetic feature $\hat{x}$ or the original feature x is applied to the input of the Decoder, and the developed synthetic feature $\hat{x}$ is applied to the hidden layer of the Generator. The enhanced synthetic feature $\hat{x}$ is again applied to the inputs of the Discriminator and Decoder. Discriminator calculates the L_W to determine whether the visual feature is real or fake. The latent embeddings transformed by the decoder are combined with the corresponding visual features x and $\hat{x}$, and trained using a cycle-consistency loss (L_R). Finally, the losses obtained from the Discriminator and Decoder are calculated for training and used for ZSL and GZSL classification.

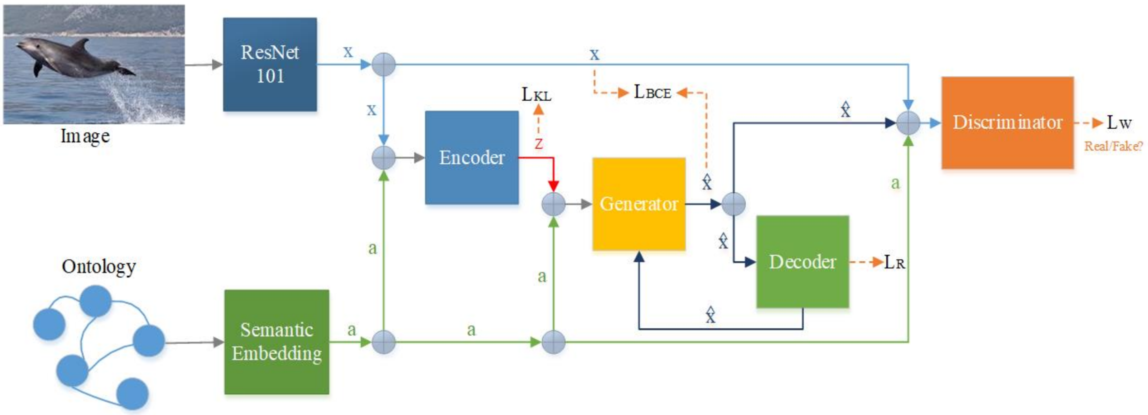

Figure 2. Illustration of the proposed architecture

4. Experimental Result

In this section, the evaluation of our proposed study, comparisons with other methods, and analyses of the model are presented.

4.1. Dataset

We evaluate our method using the AWA2 (Animals with Attributes 2) [34] dataset, which is suitable for ontology creation. The AWA2 dataset contains images from 50 classes, with 40 of these classes proposed as seen classes and 10 as unseen classes based on the Proposed Split (PS). AWA2, comprising a total of 37,322 images, is a coarse-grained dataset suitable for ontology creation. Statistical information about the AWA2 dataset is provided in Table 3.

Table 2. Statistics for the AWA2 dataset.

Dataset	Seen Classes	Unseen Classes	Total Classes	Total Images
AWA2	40	10	50	37322

4.2. Performance metrics and evaluation

In our study, we examine classification performances for both GZSL and traditional ZSL tasks. The best accuracy rate is used for ZSL performance measurements in the experiments. For GZSL performance measurements, the commonly used harmonic mean is selected. To calculate the harmonic mean, the accuracy rates of seen and unseen classes are utilized, as shown in Equation 1. In the equation, the harmonic mean (H) is defined in terms of the accuracy rates of seen classes (Acc_S) and unseen classes (Acc_U).

$$H = \frac{2 \times Acc_S \times Acc_U}{Acc_S + Acc_U} \quad (1)$$

Table 4 displays the experimental results of the proposed method and state-of-the-art methods for both ZSL and GZSL on the AWA2 dataset. The table indicates the best accuracy (T1) for ZSL performance, and for GZSL performance, it shows the results for unseen classes (U), seen classes (S), and the harmonic mean (H). Performance measurements are calculated as percentages, and the best results are highlighted in bold. Looking at Table 4, our proposed method achieved the best H value with a GZSL performance of 66.29%. For ZSL tasks, it secured the third position with a performance of 69.69%. The proposed method demonstrates a 1.19% improvement for the GZSL task.

Table 3. ZSL and GZSL experimental results performed with the AWA dataset.

Methods	**T1**	**S**	**U**	**H**
MVAAD [36].	69.50	93.40	30.70	49.20
Cramer GAN [37].	72.40	-	-	65.00
Tf-GCZSL [38].	-	64.89	40.23	48.33
DGEM [39].	67.30	69.80	43.40	53.40
SRSA [40].	68.30	59.60	38.10	46.50
Niu et al. [41].	-	66.00	59.30	62.50
ZSLGC-MLO [42].	**73.10**	71.20	57.80	63.80
Zhang et al. [43]	-	72.40	59.10	65.10
Geng et al. [26]	62.65	59.59	50.58	54.71
Ours	69.69	79.05	57.08	**66.29**

4.3. Model analysis

4.3.1. Ablation study

Ablation experiments were conducted to investigate the impact of semantic representations in our study. We evaluated different combinations of semantic attributes (att), ontology (onto), and textual descriptions (text). The results obtained from the experiments are shown in Table 5, with the best results highlighted in bold. According to the results, the best T1 and H values were achieved when using semantic attributes and ontology together.

Table 4. Ablation experiments for the AWA dataset

Methods	**T1**	**S**	**U**	**H**
Ours + att + onto	**69.69**	79.05	57.08	**66.29**
Ours + att + text	60.40	77.95	49.55	60.58
Ours + att + onto + text	67.43	73.79	60.10	66.25

4.3.2. Parameter analysis

In our experiments, we examined the effects of classification learning rates of 0.00001, 0.0001, and 0.0005 on U, S, and H. The results of the classification learning rates are shown in Fig. 3. Looking at Fig. 3, the best result was obtained with a learning rate of 0.0005. Additionally, the results indicate that different classification learning rates have an impact on the outcomes.

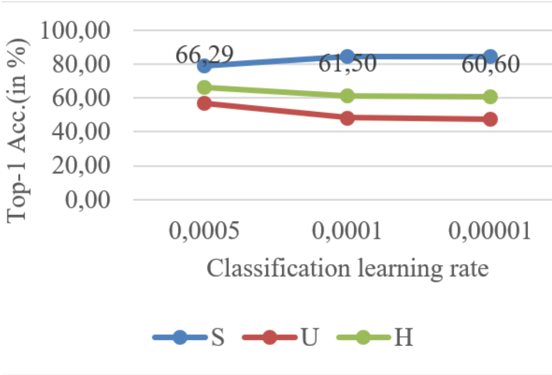

Figure 3. Results of classification learnibg rates

5. Conclusions and Discussion

In this paper, we proposed a generative-based model enriched with semantic representations for GZSL tasks. We utilized VAE-GAN generator networks to generate synthetic images for unseen classes. We leveraged an ontology created with RDF for the semantic representations of classes. Thus, by using both the numerical values of semantic attributes and their meaningful relationships, we created a highly discriminative semantic embedding. VAE establishes tight relationships between the visual and semantic domains. It is known that GAN prevents mode collapse issues. Taking advantage of the strong features of VAE and GAN, we synthesized image features and addressed data imbalance. We evaluated the performance of our proposed model on both standard ZSL and generalized ZSL using the AWA2 dataset. We compared the performance of our proposed model with the performances of the latest technology methods, achieving an additional improvement of 1.19 points for the GZSL task.

In our future work, we aim to enhance the success achieved in this study. We plan to create an ontology using the OWL language. We believe that incorporating the innovations brought by the OWL language into our ontology structure will contribute to the literature by improving the performance of our model.

Conflict of Interest Statement

The authors declare that there is no conflict of interest.

References

[1] F. Lv, J. Zhang, G. Yang, L. Feng, Y. Yu, and L. Duan, "Learning cross-domain semantic-visual relationships for transductive zero-shot learning," *Pattern Recognit.*, vol. 141, p. 109591, Sep. 2023. doi:10.1016/j.patcog.2023.109591

[2] W. Alhoshan, A. Ferrari, and L. Zhao, "Zero-shot learning for requirements classification: An exploratory study," *Inf. Softw. Technol.*, vol. 159, p. 107202, Jul. 2023. doi:10.1016/j.infsof.2023.107202

[3] E. Çelik and T. Dalyan, "Unified benchmark for zero-shot Turkish text classification," *Inf. Process. Manag.*, vol. 60, no. 3, p. 103298, May 2023. doi:10.1016/j.ipm.2023.103298

[4] "Zero-shot stance detection via multi-perspective contrastive learning with unlabeled data - ScienceDirect," [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0306457323000985. [Accessed: 07 Dec. 2023].

[5] X. Li *et al.*, "A structure-enhanced generative adversarial network for knowledge graph zero-shot relational learning," *Inf. Sci.*, vol. 629, pp. 169–183, Jun. 2023. doi:10.1016/j.ins.2023.01.113

[6] J. Eronen, M. Ptaszynski, and F. Masui, "Zero-shot cross-lingual transfer language selection using linguistic similarity," *Inf. Process. Manag.*, vol. 60, no. 3, p. 103250, May 2023. doi:10.1016/j.ipm.2022.103250

[7] X. Liu, J. Gao, X. He, L. Deng, K. Duh, and Y.-Y. Wang, "Representation Learning Using Multi-Task Deep Neural Networks for Semantic Classification and Information Retrieval," in *Proceedings of the 2015 Conference of the North American Chapter of the*

Association for Computational Linguistics: Human Language Technologies, Denver, Colorado: Association for Computational Linguistics, 2015, pp. 912–921. doi:10.3115/v1/N15-1092

[8] F. Al Kassar and F. Armetta, "Extracting Tags from Large Raw Texts Using End-to-End Memory Networks," in *Proceedings of the 2nd Workshop on Semantic Deep Learning (SemDeep-2)*, D. Gromann, T. Declerck, and G. Heigl, Eds., Montpellier, France: Association for Computational Linguistics, Sep. 2017, pp. 33–40. [Online]. Available: https://aclanthology.org/W17-7305. [Accessed: 04 Dec. 2023].

[9] G. Petrucci, C. Ghidini, and M. Rospocher, "Ontology Learning in the Deep," in *Knowledge Engineering and Knowledge Management*, vol. 10024, E. Blomqvist, P. Ciancarini, F. Poggi, and F. Vitali, Eds., in Lecture Notes in Computer Science, vol. 10024. , Cham: Springer International Publishing, 2016, pp. 480–495. doi:10.1007/978-3-319-49004-5_31

[10] D. Jurafsky and J. H. Martin, "Speech and Language Processing," [Online]. Available: https://web.stanford.edu/~jurafsky/slp3/. [Accessed: 04 Dec. 2023].

[11] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3 edition. Upper Saddle River: Pearson, 2009.

[12] "Semantic Web - W3C," [Online]. Available: https://www.w3.org/standards/semanticweb/. [Accessed: 04 Dec. 2023].

[13] J. Hendler, T. Berners-Lee, and E. Miller, "Integrating applications on the semantic web," *J. Inst. Electr. Eng. Jpn.*, vol. 122, pp. 676–680, Jan. 2002. doi:10.1541/ieejjournal.122.676

[14] F. Pourpanah *et al.*, "A Review of Generalized Zero-Shot Learning Methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–20, 2022. doi:10.1109/TPAMI.2022.3191696

[15] "LearnOpenCV," [Online]. Available: https://learnopencv.com/zero-shot-learning-an-introduction/. [Accessed: 04 Dec. 2023].

[16] C. Patrício and J. C. Neves, "Zero-shot face recognition: Improving the discriminability of visual face features using a Semantic-Guided Attention Model," *Expert Syst. Appl.*, vol. 211, p. 118635, Jan. 2023. doi:10.1016/j.eswa.2022.118635

[17] J. Wu, Y. Zhang, X. Zhao, and W. Gao, "A Generalized Zero-Shot Framework for Emotion Recognition from Body Gestures," arXiv, Oct. 20, 2020. doi:10.48550/arXiv.2010.06362

[18] R. Gao *et al.*, "Zero-VAE-GAN: Generating Unseen Features for Generalized and Transductive Zero-Shot Learning," *IEEE Trans. Image Process.*, vol. 29, pp. 3665–3680, Jan. 2020. doi:10.1109/TIP.2020.2964429

[19] S. Narayan, A. Gupta, F. S. Khan, C. G. M. Snoek, and L. Shao, "Latent Embedding Feedback and Discriminative Features for Zero-Shot Classification." arXiv, Jul. 18, 2020. doi:10.48550/arXiv.2003.07833

[20] J. Bao, D. Chen, F. Wen, H. Li, and G. Hua, "CVAE-GAN: Fine-Grained Image Generation through Asymmetric Training." arXiv, Oct. 12, 2017. doi:10.48550/arXiv.1703.10155

[21] Z. Han, Z. Fu, G. Li, and J. Yang, "Inference guided feature generation for generalized zero-shot learning," *Neurocomputing*, vol. 430, pp. 150–158, Mar. 2021. doi:10.1016/j.neucom.2020.10.080

[22] "Semantic Deep Learning," [Online]. Available: https://www.dfki.de/~declerck/semdeep-4/index.html. [Accessed: 04 Dec. 2023].

[23] H. Wang, "Semantic Deep Learning," 2015.

[24] Y. Geng *et al.*, "Generative Adversarial Zero-shot Learning via Knowledge Graphs," arXiv, Apr. 06, 2020. doi:10.48550/arXiv.2004.03109

[25] Y. Geng *et al.*, "OntoZSL: Ontology-enhanced Zero-shot Learning," arXiv.org, [Online]. Available: https://arxiv.org/abs/2102.07339v1. [Accessed: 08 Dec. 2023].

[26] Y. Geng *et al.*, "Benchmarking knowledge-driven zero-shot learning," *J. Web Semant.*, vol. 75, p. 100757, Jan. 2023. doi:10.1016/j.websem.2022.100757

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition." arXiv, Dec. 10, 2015. doi:10.48550/arXiv.1512.03385

[28] H. Zhang, H. Que, J. Ren, and Z. Wu, "Transductive semantic knowledge graph propagation for zero-shot learning," *J. Frankl. Inst.*, vol. 360, no. 17, pp. 13108–13125, Nov. 2023. doi:10.1016/j.jfranklin.2023.07.009

[29] L. Nieto-Piña and R. Johansson, "Automatically Linking Lexical Resources with Word Sense Embedding Models," in *Proceedings of the Third Workshop on Semantic Deep Learning*, L. E. Anke, D. Gromann, and T. Declerck, Eds., Santa Fe, New Mexico: Association

for Computational Linguistics, Aug. 2018, pp. 23–29. [Online]. Available: https://aclanthology.org/W18-4003. [Accessed: 04 Dec. 2023].

[30] Y. Zhou, J. Shah, and S. Schockaert, "Learning Household Task Knowledge from WikiHow Descriptions," in *Proceedings of the 5th Workshop on Semantic Deep Learning (SemDeep-5)*, L. Espinosa-Anke, T. Declerck, D. Gromann, J. Camacho-Collados, and M. T. Pilehvar, Eds., Macau, China: Association for Computational Linguistics, Aug. 2019, pp. 50–56. [Online]. Available: https://aclanthology.org/W19-5808.

[31] D. Loureiro and A. Jorge, "LIAAD at SemDeep-5 Challenge: Word-in-Context (WiC)," in *Proceedings of the 5th Workshop on Semantic Deep Learning (SemDeep-5)*, L. Espinosa-Anke, T. Declerck, D. Gromann, J. Camacho-Collados, and M. T. Pilehvar, Eds., Macau, China: Association for Computational Linguistics, Aug. 2019, pp. 1–5. [Online]. Available: https://aclanthology.org/W19-5801. [Accessed: 04 Dec. 2023].

[32] J. Park, K. Kim, W. Hwang, and D. Lee, "Concept embedding to measure semantic relatedness for biomedical information ontologies," *J. Biomed. Inform.*, vol. 94, p. 103182, Jun. 2019. doi:10.1016/j.jbi.2019.103182

[33] T. Murata *et al.*, "Predicting Relations Between RDF Entities by Deep Neural Network," Nov. 2017, pp. 343–354. doi:10.1007/978-3-319-70407-4_43

[34] G. Petrucci, M. Rospocher, and C. Ghidini, "Expressive ontology learning as neural machine translation," *J. Web Semant.*, vol. 52–53, pp. 66–82, Oct. 2018. doi:10.1016/j.websem.2018.10.002

[35] J.-R. Ruiz-Sarmiento, C. Galindo, J. Monroy, F.-A. Moreno, and J. Gonzalez-Jimenez, "Ontology-based conditional random fields for object recognition," *Knowl.-Based Syst.*, vol. 168, pp. 100–108, Mar. 2019. doi:10.1016/j.knosys.2019.01.005

[36] L. Tian *et al.*, "Multi-scale visual attention for attribute disambiguation in zero-shot learning," *Signal Process. Image Commun.*, vol. 103, p. 116614, Apr. 2022. doi:10.1016/j.image.2021.116614

[37] J. Liu, L. Fu, H. Zhang, Q. Ye, W. Yang, and L. Liu, "Learning discriminative and representative feature with cascade GAN for generalized zero-shot learning," *Knowl.-Based Syst.*, vol. 236, p. 107780, Jan. 2022. doi:10.1016/j.knosys.2021.107780

[38] C. Gautam, S. Parameswaran, A. Mishra, and S. Sundaram, "Tf-GCZSL: Task-free generalized continual zero-shot learning," *Neural Netw.*, vol. 155, pp. 487–497, Nov. 2022. doi:10.1016/j.neunet.2022.08.034

[39] J. Zhang, Y. Geng, W. Wang, W. Sun, Z. Yang, and Q. Li, "Distribution and gradient constrained embedding model for zero-shot learning with fewer seen samples," *Knowl.-Based Syst.*, vol. 251, p. 109218, Sep. 2022. doi:10.1016/j.knosys.2022.109218

[40] Y. Liu, X. Gao, J. Han, L. Liu, and L. Shao, "Zero-shot learning via a specific rank-controlled semantic autoencoder," *Pattern Recognit.*, vol. 122, p. 108237, Feb. 2022. doi:10.1016/j.patcog.2021.108237

[41] C. Niu *et al.*, "Unbiased feature generating for generalized zero-shot learning," *J. Vis. Commun. Image Represent.*, vol. 89, p. 103657, Nov. 2022. doi:10.1016/j.jvcir.2022.103657

[42] X. Xu, X. Bao, X. Lu, R. Zhang, X. Chen, and G. Lu, "An end-to-end deep generative approach with meta-learning optimization for zero-shot object classification," *Inf. Process. Manag.*, vol. 60, no. 2, p. 103233, Mar. 2023. doi:10.1016/j.ipm.2022.103233

[43] J. Zhang, S. Liao, H. Zhang, Y. Long, Z. Zhang, and L. Liu, "Data driven recurrent generative adversarial network for generalized zero shot image classification," *Inf. Sci.*, vol. 625, pp. 536–552, May 2023. doi:10.1016/j.ins.2023.01.039