Zero-shot Classification using Hyperdimensional Computing

Samuele Ruffino*†, Geethan Karunaratne*, Michael Hersche*†,
Luca Benini†, Abu Sebastian*, and Abbas Rahimi*

*IBM Research, Zürich, Switzerland ‡ETH Zürich, Zürich, Switzerland

***Abstract*—Classification based on Zero-shot Learning (ZSL) is the ability of a model to classify inputs into novel classes on which the model has not previously seen any training examples. Providing a set of attributes associated with the new class as an auxiliary descriptor is one of the favored approaches to solving this challenging task. In this work, inspired by Hyperdimensional Computing (*HDC*), we propose the use of stationary distributed binary codebooks in an attribute encoder to compactly represent a computationally simple end-to-end trainable model, which we name Hyperdimensional Computing Zero-shot Classifier (*HDC-ZSC*). It additionally consists of a trainable image encoder, and a similarity kernel. *HDC-ZSC* achieves Pareto optimal results with a 63.8% top-1 classification accuracy on the CUB-200 dataset by having only 26.6 million trainable parameters. Compared to two other state-of-the-art non-generative approaches, *HDC-ZSC* achieves 4.3% and 9.9% better accuracy, while they require more than 1.85× and 1.72× parameters compared to *HDC-ZSC*, respectively.**

***Index Terms*—Zero-shot Learning, Hyperdimensional Computing, Fine-grained Classification**

I. Introduction

Zero-shot Learning (ZSL) aims at classifying an object without having seen any instances from the same class during training [1]–[7]. To enable the recognition of previously unseen classes, in ZSL, the learner is provided with a unique descriptor that distinguishes the unseen class with respect to the classes the learner was trained on. The descriptor is typically a collection of attributes, each of which can take a finite range of values. The goal of ZSL is, having seen certain combinations of values for the set of attributes, to successfully interpolate to other value combinations corresponding to the unseen classes.

A ZSC model generally consists of an image encoder and an attribute encoder as shown in Fig. 1. At inference time, the image encoder accepts an image from an *unseen* class on which the model was not trained. In this example, a duck image is presented to the model, but no images from the duck class were provided during training. However, the model gets access to auxiliary information of all classes including the unseen ones via the attribute encoder.

There are several approaches for solving ZSL problems, which can be broadly divided as *non-generative* and *generative*. In non-generative approaches, a mapping function generalizes alignment between image and attribute descriptions [1], [8]. The generative approaches on the other hand rely on larger models capable of artificially manufacturing instances for a given unseen class descriptor thereby helping effectively turn the problem into a few-shot learning problem [9].

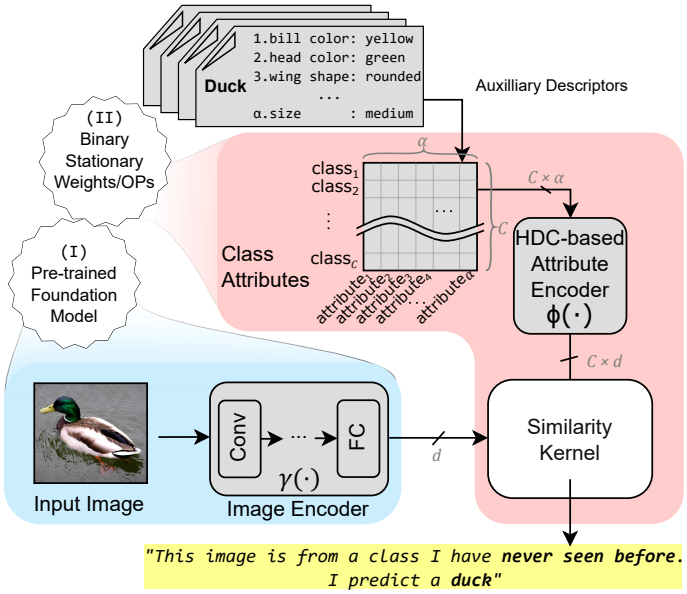

Fig. 1. The model structure of *HDC-ZSC* employed for Zero-shot Classification in this work. It comprises two main modules: a pre-trained foundation model image encoder and an HDC-based attribute encoder. The attribute encoder consists of weights with fixed binary vectors and binary vector operations, providing opportunities for implementation in resource-constrained edge devices. Modules filled with gray color remain stationary during inference.

This work presents a new non-generative end-to-end training method for Zero-shot Classification (ZSC) featuring attribute encoders containing fixed and compact codebooks whose dimensionality is chosen by the *HDC* machinery [10]. *HDC* has recently been combined with neural networks which not only achieved the state-of-the-art accuracy in image-based few-shot learning [11] and few-shot continual learning [12], [13] but also led to energy-efficient hardware implementations. Here, we further expand this combination to the ZSC tasks.

II. Method

In this work, a novel hybrid architecture for ZSL is proposed, exploiting class-attribute association and *HDC*. The concept is illustrated in Fig. 1. The inputs to the model are provided in two modalities: (i) a batch of images $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_B)$ in matrix form with batch size B, and (ii) a class attributes matrix $\mathbf{A} = (\mathbf{a}_1, ..., \mathbf{a}_C) \in \mathbb{R}^{C \times \alpha}$, where each row is an attribute vector of α dimension with C vectors in total representing all classes. The images of dimensions h, w, l are passed through

979-8-3503-4859-0/DATE24/© 2024 EDAA

an image encoder $\gamma(\cdot) : \mathbb{R}^{h\times w\times l} \rightarrow \mathbb{R}^d$ and the attributes are passed through an attribute encoder $\phi(\cdot) : \mathbb{R}^{\alpha} \rightarrow \mathbb{R}^d$, both of which generate embeddings of the same dimensionality (d). The embeddings are then compared via a bi-similarity kernel *cossim*, that relates image and class embeddings [1], [8], as follows:

$$cossim(\gamma(\mathbf{X}), \phi(\mathbf{A})) = \frac{1}{K} \frac{\gamma(\mathbf{X})^T \cdot \phi(\mathbf{A})}{||\gamma(\mathbf{X})||\ ||\phi(\mathbf{A})||},$$

where K is the learnable temperature scaling parameter. The resulting similarity matrix of dimension $B \times C$ is used for error computation with respect to the batch of ground-truth labels and updating the image encoder's weights over a series of epochs. Finally, the predicted class for every unseen queried image $\mathbf{x}$ is computed as

$$\hat{\mathbf{y}} = \underset{i\in\{1,\ldots,C\}}{\arg\max}\ cossim(\gamma(\mathbf{x}), \phi(\mathbf{a}_i)).$$

We also propose a training methodology to boost ZSC accuracy beyond previous state-of-the-art non-generative methods. It consists of pre-training the model on a standard image classification task at first, followed by a domain-specific attribute extraction task. Here we predict the attributes present in an image to match the ground-truth attributes of the target. In the final phase of training, the matured model is exposed to the ZSC task by retraining to discriminate attribute vectors representing the classes.

III. Evaluation

For evaluation, we use Caltech-UCSD Birds-200-2011 (CUB-200) [14] as the dataset and top-1 accuracy as the metric.

Fig. 2 compares *HDC-ZSC* accuracy performance and number of parameters with other state-of-the-art approaches, both generative and non-generative methods. Additionally, results from an alternative *Trainable-MLP model*, in which fixed *HDC* codebooks-based attribute encoders are replaced by 2-layer trainable MLP, are included in the table for reference. When compared to the state-of-the-art generative methods, our models reduce the number of total network parameters by more than 1.75×–2.58× while maintaining a competitive top-1% accuracy (within 3.9%). Specifically, comparing *HDC-ZSC* with ESZSL [1], as a non-generative model similar to our approach, we obtain a $+9.9\%$ accuracy improvement while improving the parameter efficiency by more than 1.72×.

When compared with both state-of-the-art generative and non-generative approaches, *HDC-ZSC*, and its variants, are in the Pareto front with respect to accuracy and model parameter count (see Fig. 2). This could potentially pave the way for its implementation as an application target in low-power embedded platforms [15].

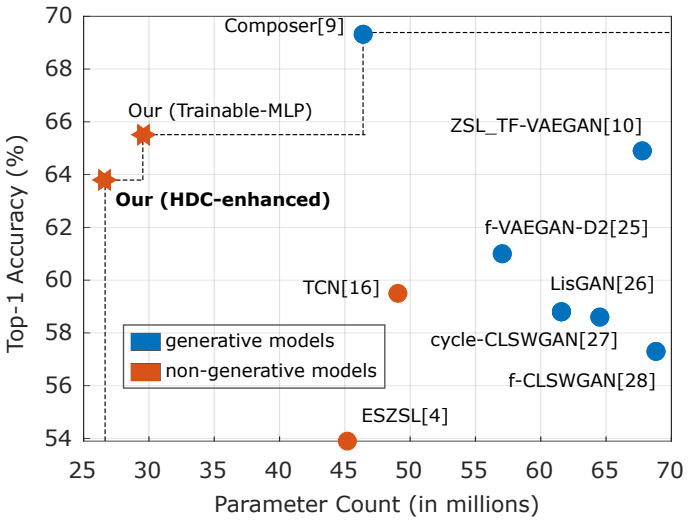

Fig. 2. Comparison of Zero-shot Classification accuracy vs model parameter count. Our models *HDC-ZSC* and *Trainable-MLP model* are both in the Pareto front.

Acknowledgments

We would like to acknowledge the early technical investigations by Junyuan Cui.

References

[1] B. Romera-Paredes and P. Torr, "An embarrassingly simple approach to zero-shot learning," in *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, vol. 37. PMLR, 2015, pp. 2152–2161.

[2] E. Altszyler, P. Brusco, N. Basiou *et al.*, "Zero-shot multi-domain dialog state tracking using descriptive rules," *International Workshop on Neural-Symbolic Learning and Reasoning (NeSy)*, 2020.

[3] M. Norouzi, T. Mikolov, S. Bengio *et al.*, "Zero-shot learning by convex combination of semantic embeddings," *arXiv preprint arXiv:1312.5650*, 2013.

[4] E. Kodirov, T. Xiang, and S. Gong, "Semantic autoencoder for zero-shot learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3174–3183.

[5] Z. Zhang and V. Saligrama, "Zero-shot learning via semantic similarity embedding," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[6] D. Huynh and E. Elhamifar, "Compositional fine-grained low-shot learning," *arXiv preprint arXiv:2105.10438*, 2021.

[7] S. Narayan, A. Gupta, F. S. Khan *et al.*, "Latent embedding feedback and discriminative features for zero-shot classification," in *European Conference on Computer Vision (ECCV)*. Springer, 2020, pp. 479–495.

[8] A. Radford, J. W. Kim, C. Hallacy *et al.*, "Learning transferable visual models from natural language supervision," in *International Conference on Machine Learning (ICML)*. PMLR, 2021, pp. 8748–8763.

[9] J. Li, M. Jing, K. Lu *et al.*, "Leveraging the invariant side of generative zero-shot learning," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 7394–7403.

[10] P. Kanerva, "Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors," *Cognitive Computation*, vol. 1, no. 2, pp. 139–159, 2009.

[11] G. Karunaratne, M. Schmuck, M. Le Gallo *et al.*, "Robust high-dimensional memory-augmented neural networks," *Nature Communications*, vol. 12, no. 1, pp. 1–12, 2021.

[12] M. Hersche, G. Karunaratne, G. Cherubini *et al.*, "Constrained Few-shot Class-incremental Learning," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[13] G. Karunaratne, M. Hersche, J. Langeneager *et al.*, "In-memory realization of in-situ few-shot continual learning with a dynamically evolving explicit memory," in *IEEE 48th European Solid State Circuits Conference (ESSCIRC)*, 2022, pp. 105–108.

[14] C. Wah, S. Branson, P. Welinder *et al.*, "Cub-200 2011 dataset," California Institute of Technology, Tech. Rep. CNS-TR-2011-001, 2011.

[15] P. D. Schiavone, D. Rossi, A. Pullini *et al.*, "Quentin: an ultra-low-power pulpissimo soc in 22nm fdx," in *2018 IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S)*, 2018, pp. 1–3.