SMU SINGAPORE MANAGEMENT UNIVERSITY

Institute for Infocomm Research I2R

# Learning to Identify Seen, Unseen and Unknown in the Open World: A Practical Setting for Zero-Shot Learning

Sethupathy Parameswaran, Yuan Fang, Chandan Gautam, Savitha Ramasamy, Xiaoli Li

## Motivation

### Problem

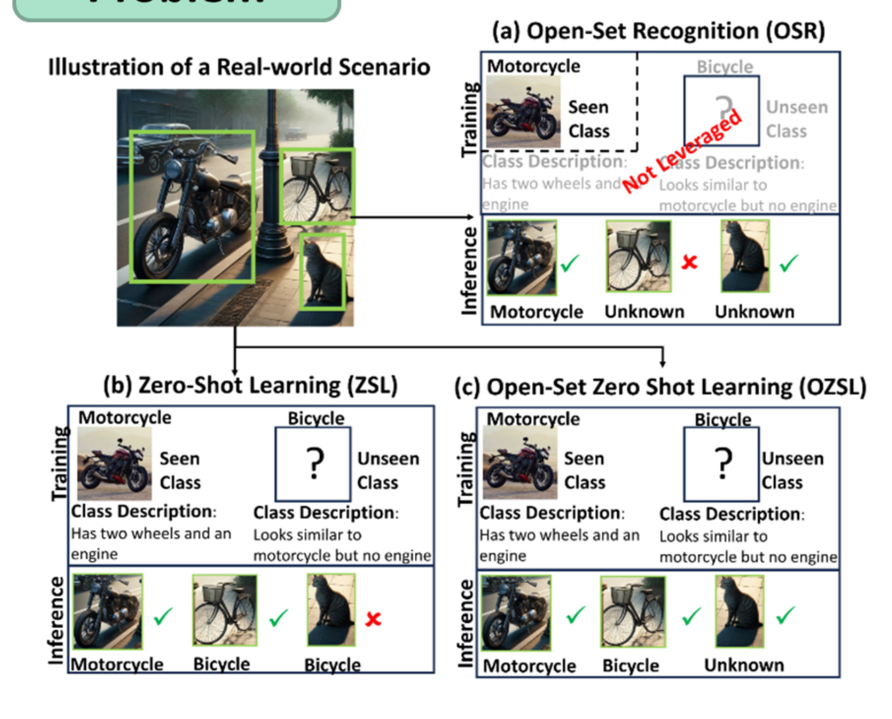


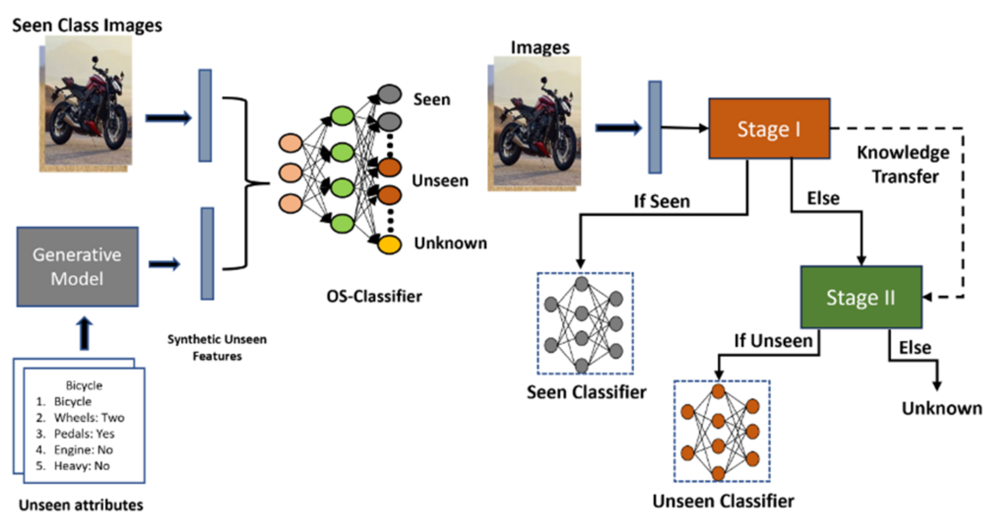

Figure 1. Illustration[1] of different classification settings. (a) OSR methods can flag unknown samples but cannot recognize unseen class samples, where class descriptions for unseen classes are not utilized. (b) ZSL methods can identify unseen class samples but cannot flag unknown samples. (c) Our OZSL approach aims to identify seen and unseen class samples, and at the same time flag unknown samples.

- ZSL methods [1,2] learn from seen classes and generalize to unseen classes by just leveraging class description.
- However, in practical scenarios, there is no guarantee that the model will only encounter classes it is trained to identify
- Hence, it must also be able to flag these samples as unknown

### Challenges

1. OSR Methods [3,4] require visual information of in-distribution (ID) samples for learning to distinguish ID samples from unknown samples.
2. However, in the proposed OZSL setting, a category of in-distribution samples namely, Unseen Classes, do not have visual information.

## One/Two Stage Approaches




(a) One-stage approach (b) Two-stage approach

Figure 2. Illustration of one-stage and two-stage approaches for open-set zero-shot learning.

## The proposed model: OZSL

### Overall framework

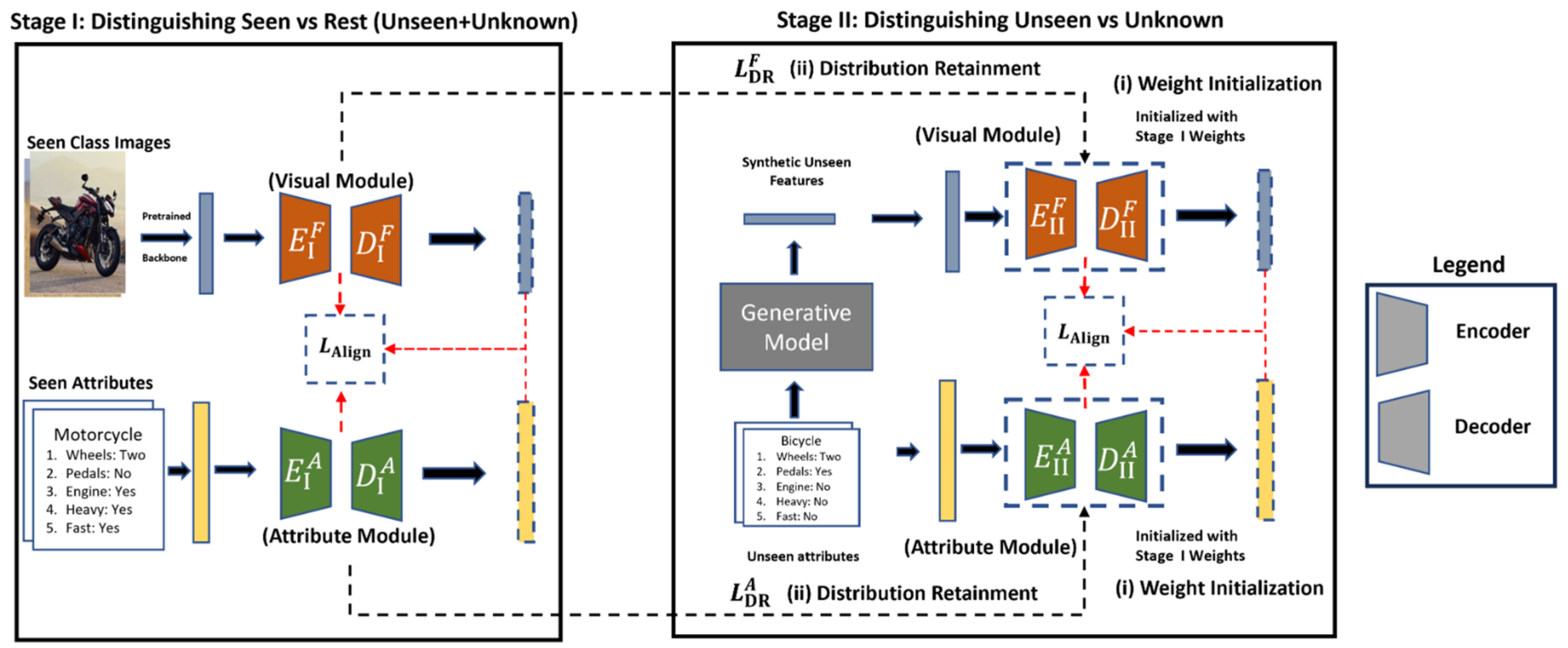



Figure 3. Illustration of the training process of our proposed method. Stage I is trained using the training data of seen classes to distinguish seen class samples from the rest. Stage II is trained using synthetic unseen class samples to distinguish unseen class samples from the unknowns. To leverage the semantic relatedness between seen and unseen classes, we propose two strategies for cross-stage knowledge transfer: (i) Weight Initialization and (ii) Distribution Retainment.

### Training: Stage I

$$L_{\text{Align}} = L^F_{\text{VAE}} + L^A_{\text{VAE}} + \lambda_{\text{cr}} L_{\text{CR}} + \lambda_{\text{cls}} L_{\text{cls}}$$

### Training: Stage II

**Weight Initialization:**

$$\Theta(E^{II}_F, D^{II}_F, E^{II}_A, D^{II}_A) \leftarrow \Theta^*(E^I_F, D^I_F, E^I_A, D^I_A)$$

**Distribution Retainment:**

$$L^F_{\text{DR}} = \|\phi^F - \phi^{F^*}\|^2_2 \qquad L^A_{\text{DR}} = \|\phi^A - \phi^{A^*}\|^2_2$$

**Overall Loss**

$$L^{\text{II}} = L_{\text{Align}} + \lambda_{\text{DR}} L_{\text{DR}}$$

### Inference

$$\omega^{\text{I}} = \begin{cases} \texttt{seen}, & \max_{\mathbf{a}\in\mathbf{A}^s}(\cos(\mathbf{z}_{\mathbf{x}^{ts}}, \mathbf{z}_{\mathbf{a}})) \geq \gamma^{\text{I}} \\ \texttt{rest}, & \text{otherwise} \end{cases}$$

$$\omega^{\text{II}} = \begin{cases} \texttt{unseen}, & \max_{\mathbf{a}\in\mathbf{A}^u}(\cos(\mathbf{z}_{\mathbf{x}^{ts}}, \mathbf{z}_{a})) \geq \gamma^{\text{II}} \\ \texttt{unknown}, & \text{otherwise} \end{cases}$$

## Experiments

### Datasets

| Dataset | AWA1 | CUB | FLO | SUN |
|---|---|---|---|---|
| # Images | 30,475 | 11,788 | 8,189 | 14,340 |
| Attribute length | 85 | 1,024 | 1,024 | 102 |
| Seen classes | 40 | 150 | 82 | 645 |
| Unseen classes | 5 | 25 | 10 | 36 |

Table 1. Statistics of the datasets.

- We use the common ZSL benchmark datasets for evaluating the performance of OZSL methods
- ZSL datasets provide a standard split for the seen and unseen classes
- We maintain the same split for the seen classes
- For the unseen classes, we randomly select half of the classes in the standard split for unseen classes
- The remaining unseen classes in the standard split are taken as the unknown class

### Main results

| | AWA1 | | | | CUB | | | | FLO | | | | SUN | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method | Seen | Unseen | Unk | HM | Seen | Unseen | Unk | HM | Seen | Unseen | Unk | HM | Seen | Unseen | Unk | HM |
| GSM (ZSL) | 79.9±2.5 | 82.3±5.6 | - | - | 63.7±1.1 | 71.4±3.7 | - | - | 88.7±0.6 | 88.2±5.1 | - | - | 31.5±0.5 | 37.7±1.1 | - | - |
| CLIP (ZSL) | / | / | / | / | 60.5±0.4 | 61.5±5.7 | - | - | 71.3±0.2 | 81.1±9.3 | - | - | 53.7±0.2 | 59.2±2.8 | - | - |
| MSP (OSR) | 74.1±0.2 | - | 89.5±5.7 | - | 59.9±0.3 | - | 49.4±3.3 | - | 78.6±0.6 | - | 61.7±10.3 | - | 34.8±0.1 | - | 61.4±1.5 | - |
| ViM (OSR) | 26.6±2.3 | - | 40.8±11.0 | - | 43.3±1.2 | - | 44.4±4.8 | - | 51.9±1.5 | - | 23.1±6.6 | - | 34.7±0.2 | - | 15.1±1.4 | - |
| KNN (OSR) | 44.3±2.8 | - | 72.3±11.2 | - | 39.7±0.9 | - | 31.1±3.6 | - | 58.4±1.4 | - | 35.1±6.8 | - | 38.1±0.3 | - | 24.9±1.7 | - |
| GSM-MSP (OZSL) | 60.4±1.5 | 43.1±10.7 | 83.3±2.4 | 57.8±5.5 | 41.4±1.8 | 39.1±3.9 | 70.7±3.8 | 45.3±1.4 | 30.2±4.2 | 33.4±9.1 | 93.6±3.6 | 36.8±4.9 | 29.1±0.2 | 36.9±1.2 | 15.9±1.1 | 25.2±0.7 |
| GSM-ViM (OZSL) | 19.1±1.4 | 42.5±4.1 | 24.2±7.1 | 23.1±1.6 | 41.1±1.2 | 41.3±3.2 | 43.9±5.1 | 41.6±1.6 | 51.1±0.7 | 67.5±5.3 | 20.9±1.8 | 36.4±1.8 | 26.3±0.5 | 18.1±0.6 | 15.1±0.7 | 20.2±0.6 |
| GSM-KNN (OZSL) | 47.3±1.1 | 47.2±10.3 | 65.2±9.4 | 50.3±4.8 | 36.4±2.1 | 45.1±3.1 | 36.2±4.9 | 38.1±1.7 | 55.3±1.3 | 54.4±3.1 | 32.8±3.6 | 44.7±2.5 | 26.7±0.4 | 28.1±0.9 | 23.4±0.1 | 26.1±0.4 |
| CLIP-MSP (OZSL) | / | / | / | / | 57.1±0.6 | 57.7±6.6 | 30.4±3.6 | 46.7±2.8 | 61.8±0.5 | 73.5±10.1 | 54.8±7.3 | 61.9±2.2 | 53.4±0.2 | 58.7±2.8 | 10.1±1.2 | 26.1±1.8 |
| CLIPN (OZSL) | / | / | / | / | 61.7±0.8 | 57.7±7.5 | 13.7±2.5 | 32.2±3.1 | 69.2±0.2 | 75.1±8.3 | 20.6±9.2 | 42.4±8.6 | 54.7±0.1 | 60.9±3.4 | 10.6±2.8 | 26.8±4.3 |
| Ours (OZSL) | 69.5±1.0 | 47.9±3.7 | 51.6±5.4 | **57.9**±3.1 | 46.1±0.3 | 41.9±5.4 | 53.8±2.2 | **47.5**±2.4 | 80.1±0.6 | 51.1±9.1 | 61.1±7.1 | **65.5**±5.2 | 29.3±2.1 | 31.7±1.1 | 40.8±2.2 | **32.1**±1.3 |

ZSL methods cannot identify Unknown Samples

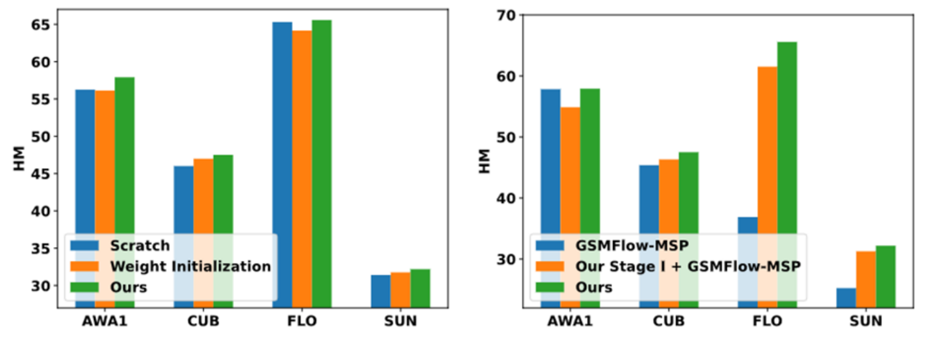
OSR methods cannot identify Unseen Samples

One-stage methods

Proposed two-stage method

Table 2. Top-1 accuracy of our proposed OZSL method and the baselines. '-' denotes that the method cannot handle a certain category of samples. '/' denotes that CLIP-based methods cannot be applied to the AWA1 dataset, as image input is required by CLIP but AWA1 has not made the images public. GSMFlow is abbreviated as 'GSM'.

### Ablation study




(a) Effect of knowledge transfer (b) Effect of two-stage learning

Figure 4. Results of ablation studies.

### Sensitivity Analysis

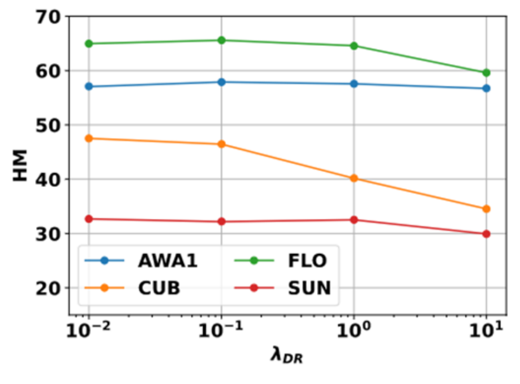
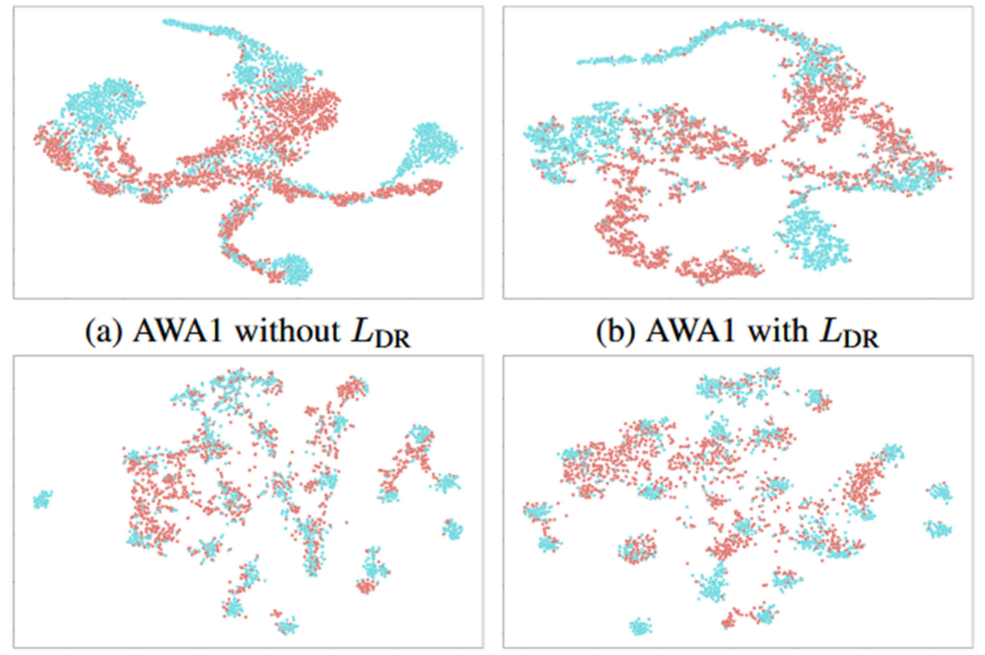



Figure 5. Effect of $\lambda_{\text{DR}}$ on the overall model performance.

### Visualization



(a) AWA1 without $L_{\text{DR}}$ (b) AWA1 with $L_{\text{DR}}$

(c) CUB without $L_{\text{DR}}$ (d) CUB with $L_{\text{DR}}$

Figure 6. *t*-SNE visualization of the latent space for Stage II of our proposed model, without or with the distribution retainment loss $L_{\text{DR}}$. The blue points represent the unseen samples and the red points represent the unknown samples.

## Conclusion

- Detection of unintentionally encountered unknown samples during inference in the real world is an important problem that needs to be addressed to ensure the trustworthiness of the mode.
- We tackle a novel and practical problem, i.e., OZSL by proposing a two-stage approach wherein we first identify the seen class samples from the rest (unseen and unknown) in Stage I and identify unseen class samples from unknown samples in Stage II
- Furthermore, we propose a cross-stage knowledge transfer mechanism in order to leverage the semantic relatedness between seen and unseen classes, to improve the overall performance of the model

## References

[1] Chen, Zhi, et al. "Mitigating generation shifts for generalized zero-shot learning." ACM MM 2021

[2] Chen, Xingyu, et al. "A Boundary Based Out-of-Distribution Classifier for Generalized Zero-Shot Learning." ECCV 2020

[3] Wang, Haoqi, et al. "Vim: Out-of-distribution with virtual-logit matching." CVPR 2022

[4] Bendale, Abhijit, and Terrance E. Boult. "Towards open set deep networks." CVPR 2016