ICCKE
2022

Averting Mode Collapse for Generative Zero-Shot Learning

Presenter : Shayan Ramazi

Authors :

Shayan Ramazi

Setare Shabani

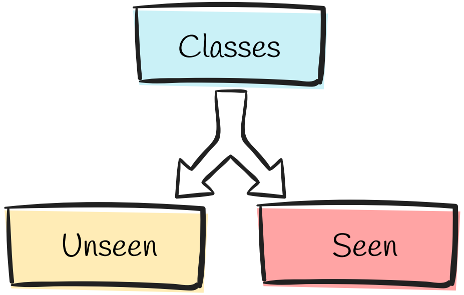
Zero Shot Learning...

- **Standard supervised learning** methods require **labeled data** in the training phase. (What happens if we do not have labeled data?)
- **Zero-Shot Learning (ZSL)** is an emerging learning paradigm that addresses the problem of recognizing **unseen classes** during training.
- Teach computers to recognize something **they have not seen.**

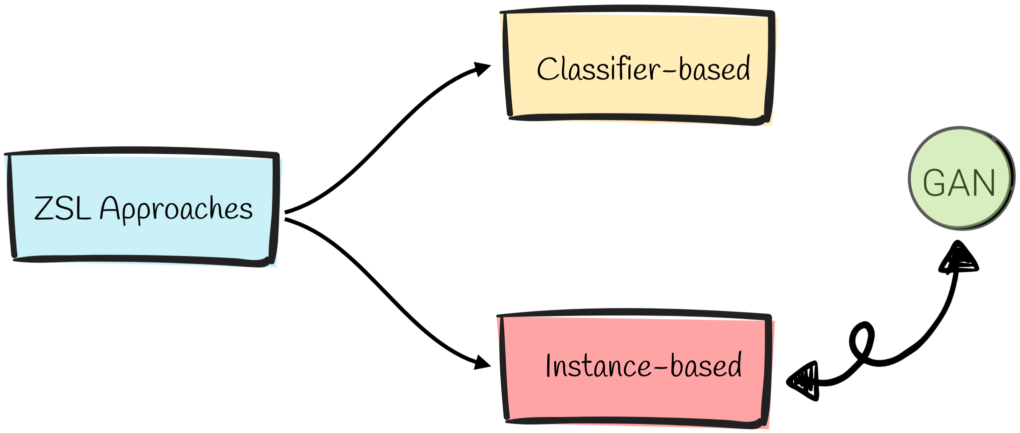
Introduction

Zero Shot Learning...

- $\{ (x_i , a_i , y_i) \in C_i \mid x \in X , a \in A , y \in Y$ **and** $i=\{S$ **or** $U\} \}$
- $\{ (x_i , a_i) \mid i=U \} \longrightarrow \{ y_i \mid i=U\}$ ◀ ZSL
- $\{ (x_i , a_i) \mid i=\{U \cup S\} \} \longrightarrow \{ y_i \mid i=\{U \cup S\}\}$ ◀ GZSL



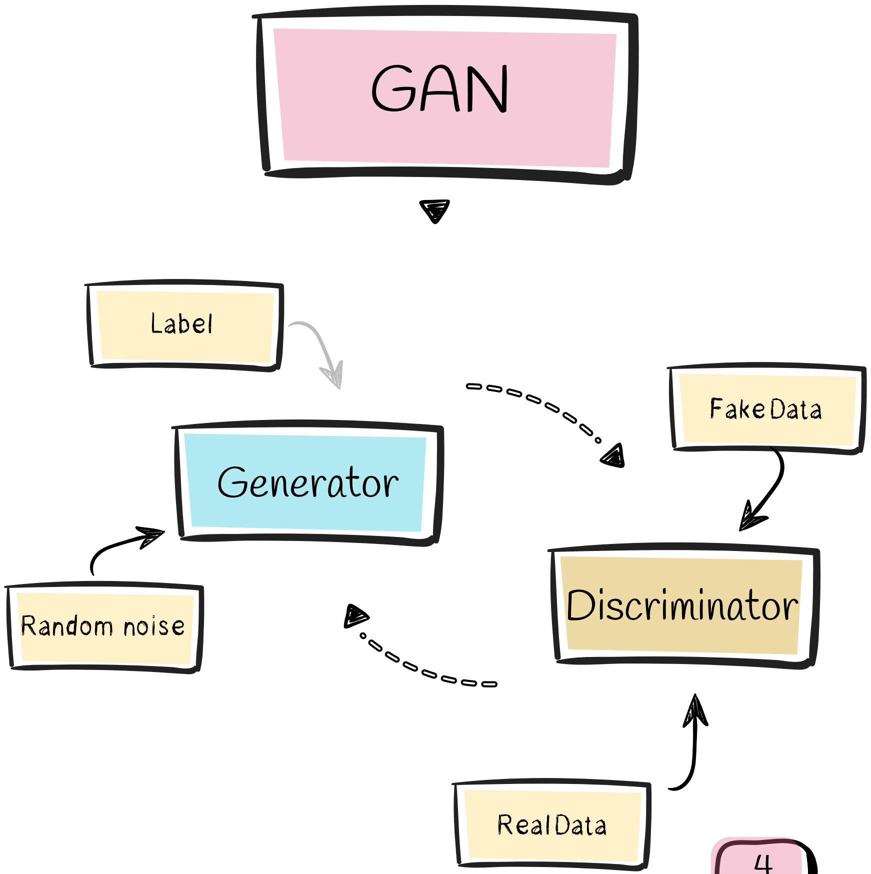
Zero Shot Learning...



GANs and Conditional GANs...

- **Generative adversarial network** (GAN) consists of a **Generator** and a **Discriminator**. The Generator's purpose is to generate fake data so that the Discriminator can not distinguish whether it is real or fake. The purpose of the Discriminator is to detect fake data from real data.

- WGAN Vs. GAN



Generative ZSL...

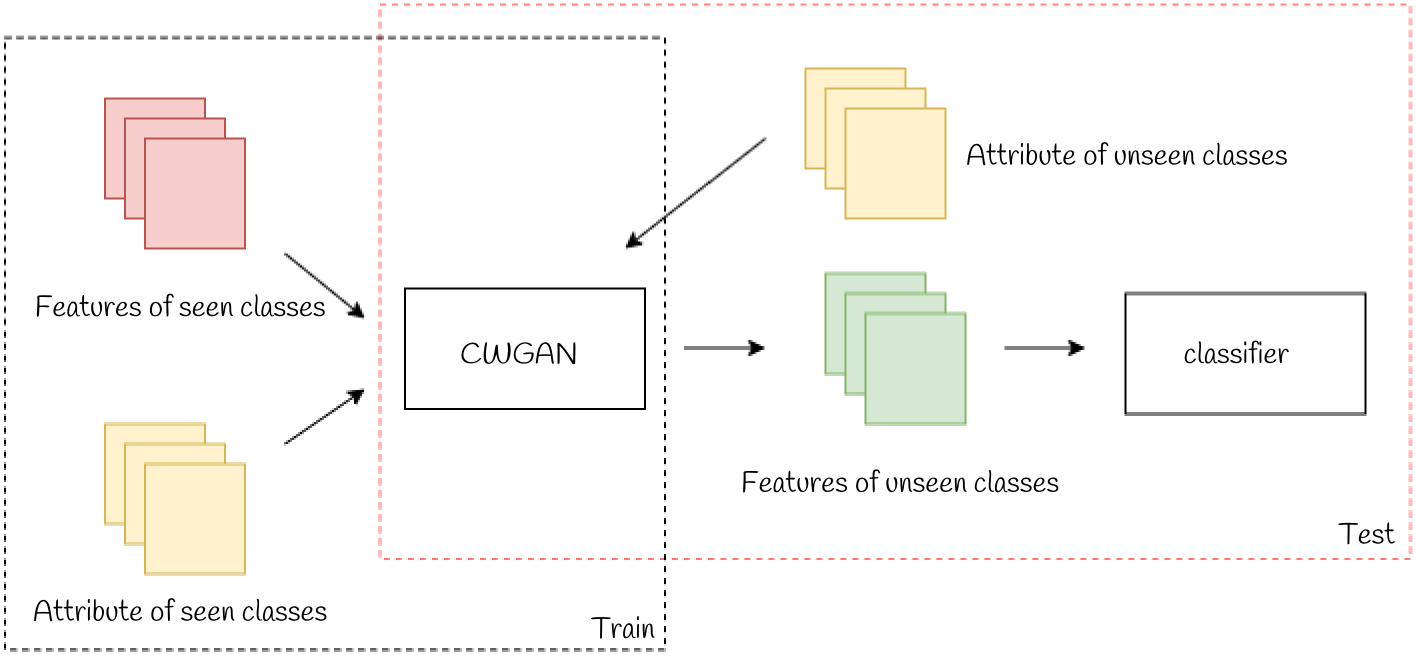


Fig. 1. General overview of Generative ZSL

Ranking Synthetic Features...

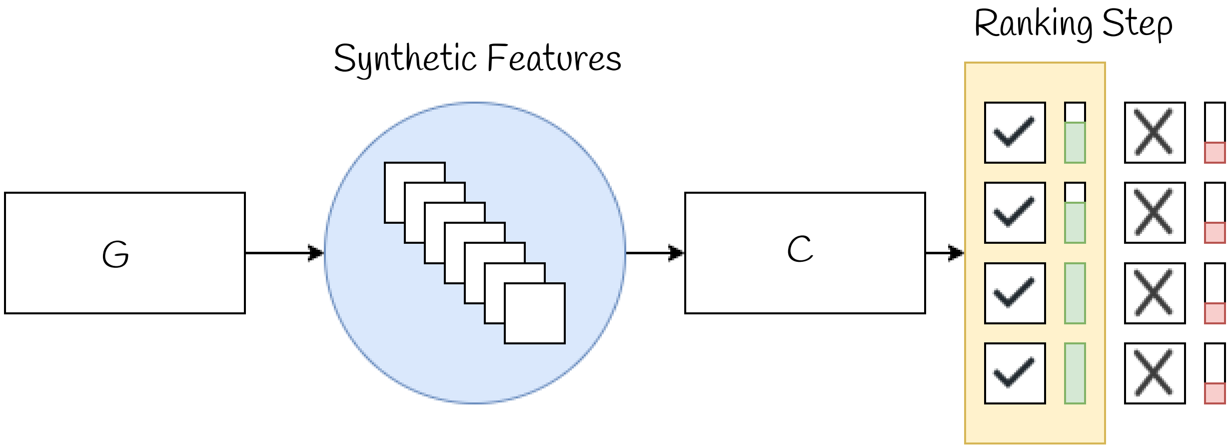


Fig. 2. Ranking Features of WGAN

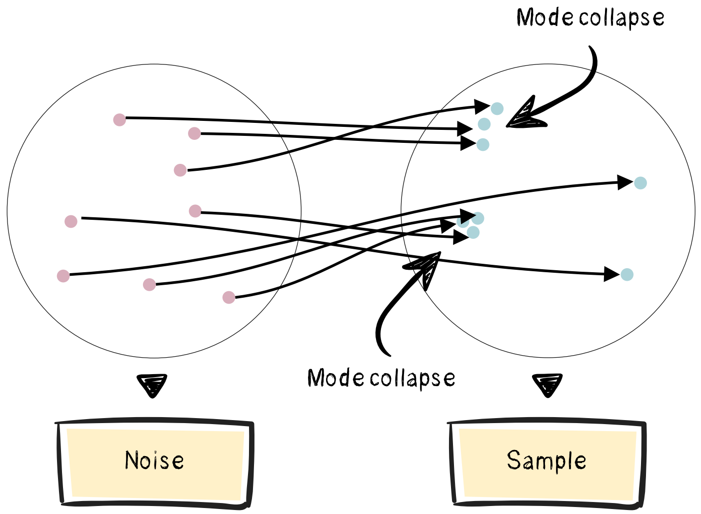
$$\theta_G = \theta_G - \alpha_G(\sum_{\gamma} \nabla_{\theta_G} v)$$

S. Ramazi and A. Nadian-Ghomsheh, "**Ranking Synthetic Features for Generative Zero-Shot Learning**," in **2021** 26th International Computer Conference, Computer Society of Iran (CSICC), 2021: IEEE, pp. 1-5.

Mode Collapse...

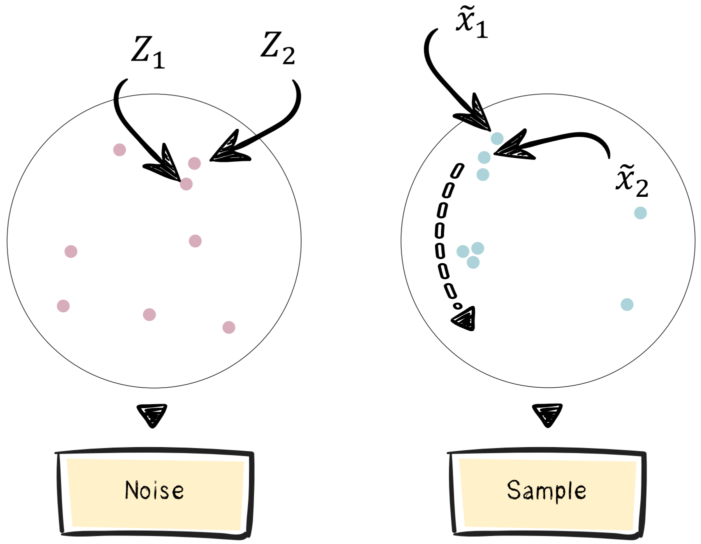
Usually we want our GAN to produce a wide variety of outputs. We want, for example, a different face for every random input to our face generator. However, if a generator produces an especially plausible output, the generator may learn to produce only that output. In fact, the generator is always trying to find the one output that seems most plausible to the discriminator.

If there is little diversity in the output and some of them are almost identical, then there is likely mode collapse.



Anti-collapse regularization term…

Remember that we sample two random variables Z_1 and Z_2. We generate two fake feature vectors $\tilde{x}_1$ and $\tilde{x}_2$ from them. When Z_1 and Z_2 are closer, $\tilde{x}_1$ and $\tilde{x}_2$ are more likely to be collapsed into the same mode. To mitigate this, the anti-collapse is defined regularization term as below. We can observe that this term amplifies the dissimilarity of the two fake feature vectors when the latent codes generating them are of high similarity.



$$\mathcal{L}_{ant} = \mathbb{E}\left[\frac{1-\cos(\tilde{x}_1,\tilde{x}_2)}{1-\cos(Z_1,Z_2)}\right],$$

Feature Generating Networks for ZSL...

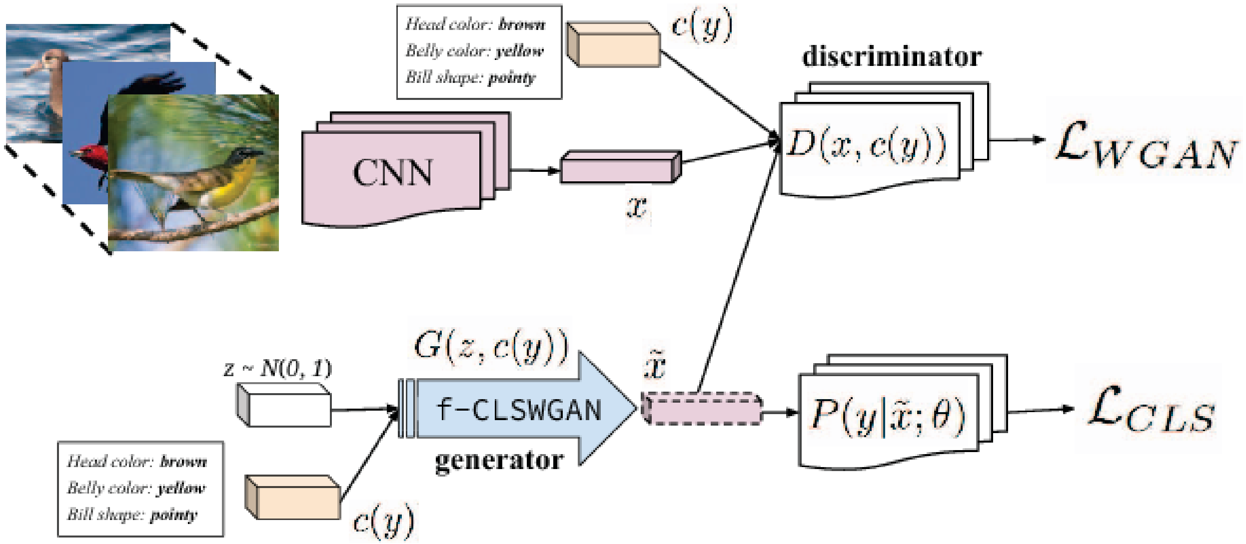


Fig. 4. f-CLSGAN

Y. Xian, T. Lorenz, B. Schiele, and Z. Akata, "**Feature generating networks for zero-shot learning,**" in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 5542-5551.classes.

Averting Mode Collapse for Generative Zero-Shot Learning...

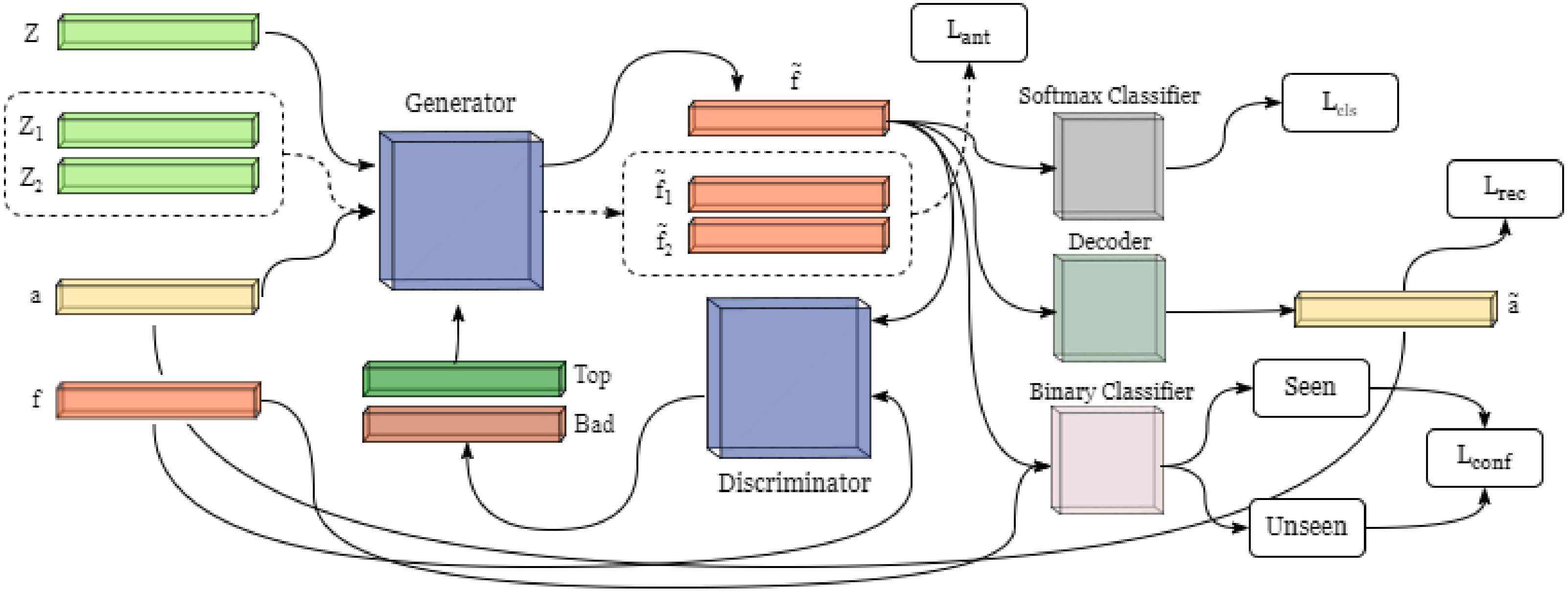
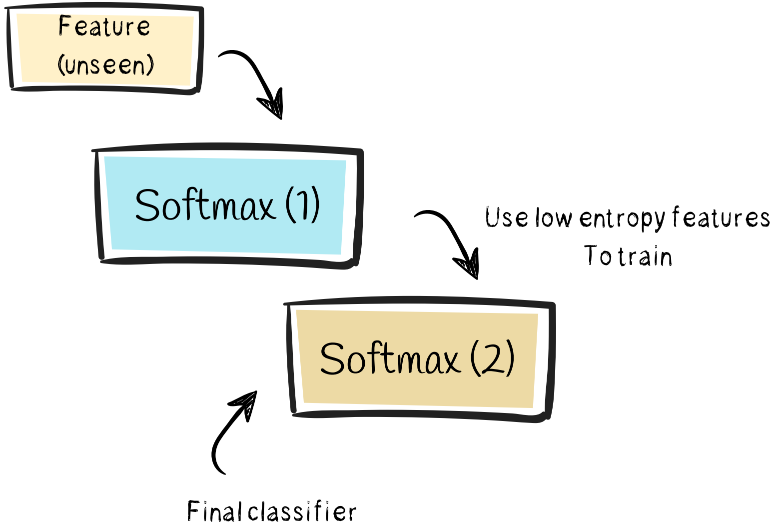


Fig. 4. General overview of our method

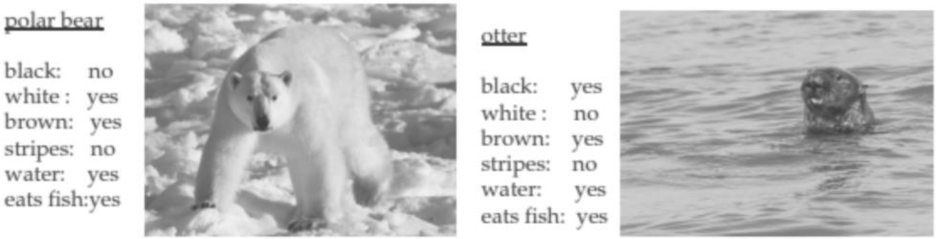
Classification...

this research employs a cascaded classifier. In the first layer of the classifier, we use a softmax classifier. The output of this layer is the possibility of assigning a sample to all classes, so it can be concluded from the output of this layer which samples are more reliable. Outputs with less classification entropy are leverage as reliable samples to train the next layer, a softmax classifier.



$$ent = -\sum_{c=1}^{number\ of\ all\ classes} y_c - log(y_c),$$

Dataset (Animal With Attribute)...



Attributes	Unseen	Seen	All	Number of Samples
85	10	40	50	30475

Table 1. Dataset

C. H. Lampert, H. Nickisch, and S. Harmeling, "**Learning to detect unseen object classes by between-class attribute transfer**," in **2009** IEEE Conference on Computer Vision and Pattern Recognition, 2009: IEEE, pp. 951-958.

Results …

In the ZSL configuration, the arithmetic mean of the top-one accuracy scores in each class is considered. Under GZSL settings, the average of the top-1 accuracy is provided by u and s (for the unseen and seen classes respectively), and the harmonic average is calculated.

$$H = ((s \times u) \times 2) / (s + u)$$

Results...

Methods	*ZSL*	*GZSL*		
		Seen	*Unseen*	*Mean*
DAP	44.1	88.7	0.0	0.0
GAZSL	68.2	86.5	19.2	31.4
f-CLSWGAN	68.2	61.4	57.9	59.6
AFC-GAN	69.1	66.8	58.2	62.2
LisGAN	70.6	76.3	52.6	62.3
Ranking Features	72.9	77.4	52.35	62.5
Ours	**74.6**	**73.9**	**57.7**	**64.8**

Table 2. Accuracy

Model analysis...

Setting	Description	Accuracy	Setting	Description
S1	WGAN+ softmax classifier	68.2	S1	WGAN+ softmax classifier
S2	Ranking Features + softmax classifier	68.5	S2	Ranking Features + softmax classifier
S3	Ranking Features + anti mode collapse loss + softmax classifier	70.3	S3	Ranking Features + anti mode collapse loss + softmax classifier
S4	Ranking Features + anti mode collapse loss + cascade classifier	74.6	S4	Ranking Features + anti mode collapse loss + cascade classifier

Table 3. Model analysis

THANKS!