Generalized Zero- and Few-Shot Learning via Aligned Variational Autoencoders

Reporter：陈思玉

2023.11.18

Schonfeld E, Ebrahimi S, Sinha S, et al. Generalized zero-and few-shot learning via aligned variational autoencoders[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 8247-8255.

Author

CENTRAL SOUTH UNIVERSITY 中南大學

Edgar Schönfeld

关注

其他姓名

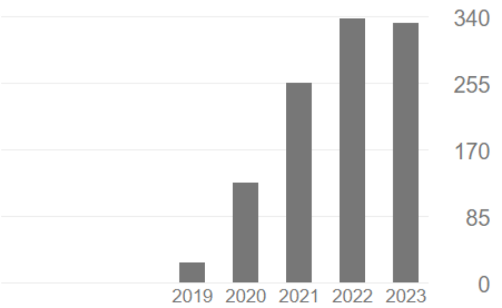
Meta
在 meta.com 的电子邮件经过验证 - 首页
Deep Learning　Computer Vision

标题	引用次数	年份
Generalized zero-and few-shot learning via aligned variational autoencoders E Schönfeld, S Ebrahimi, S Sinha, T Darrell, Z Akata Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern …	611	2019
A U-Net Based Discriminator for Generative Adversarial Networks E Schönfeld, B Schiele, A Khoreva Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern …	275	2020
You Only Need Adversarial Supervision for Semantic Image Synthesis E Schönfeld, V Sushko, D Zhang, J Gall, B Schiele, A Khoreva International Conference on Learning Representations	147 *	2021
Ex vivo treatment of prostate tumor tissue recapitulates in vivo therapy response W Zhang, WM van Weerden, CMA de Ridder, S Erkens‐Schulze, ... SCI基础版 医学3区　IF 2.8　SWJTU A+ The Prostate 79 (4), 390-402	37	2019
OASIS: only adversarial supervision for semantic image synthesis V Sushko, E Schönfeld, D Zhang, J Gall, B Schiele, A Khoreva EI检索　SCI基础版 工程技术2区　IF 19.5　SWJTU A++　SWUFE A International Journal of Computer Vision 130 (12), 2903-2923	14	2022
Training a generator neural network using a discriminator with localized distinguishing information A Khoreva, E Schoenfeld US Patent App. 17/138,748	1	2021
Cross-Linked Variational Autoencoders for Generalized Zero-Shot Learning E Schönfeld, S Ebrahimi, S Sinha, T Darrell, Z Akata	1	2019

创建我的个人资料

引用次数

	总计	2018 年至今
引用	1086	1086
h 指数	5	5
i10 指数	5	5



开放获取的出版物数量

查看全部

0 篇文章　2 篇文章

无法查看的文章　可查看的文章

根据资助方的强制性开放获取政策

合著作者

Zero-Shot Learning

ZSL 目标

ZSL 旨在训练个模型，该模型能够通过**语义信息**的辅助，利用从 seen classes 中学到的知识来对 unseen classes 进行分类。

ZSL 所用数据

- seen classes：X^s（图像特征），Y^s（类别标签），A^s（语义信息）
- unseen classes：A^u（语义信息）

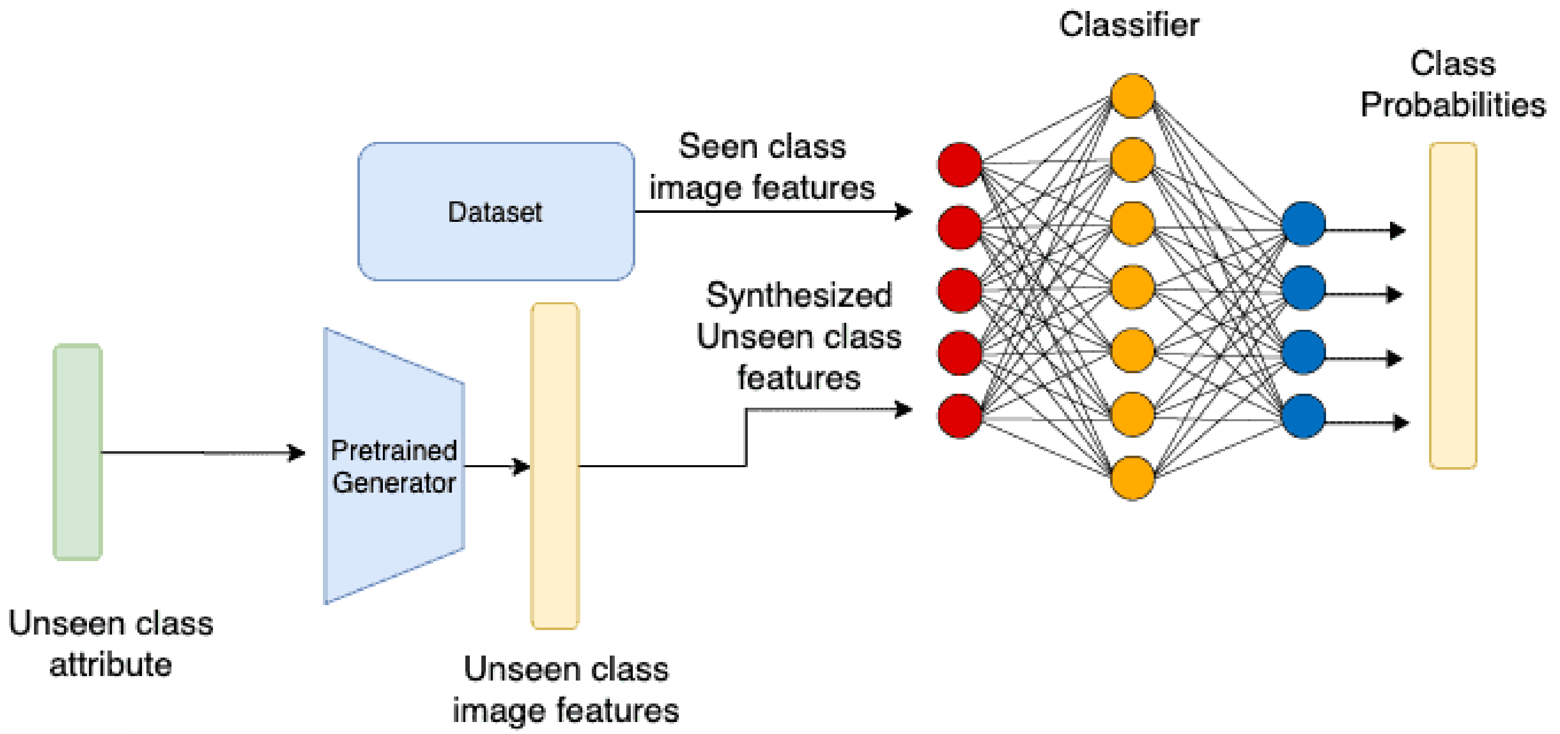
举例说明

1. 训练集有马、老虎、熊猫的图片
2. 语义信息有形状、条纹、颜色等属性
3. 给出斑马的定义：马的形状、老虎的条纹、熊猫的颜色
4. 输入斑马的图像，分类器能输出斑马的类别

Generative-based Methods

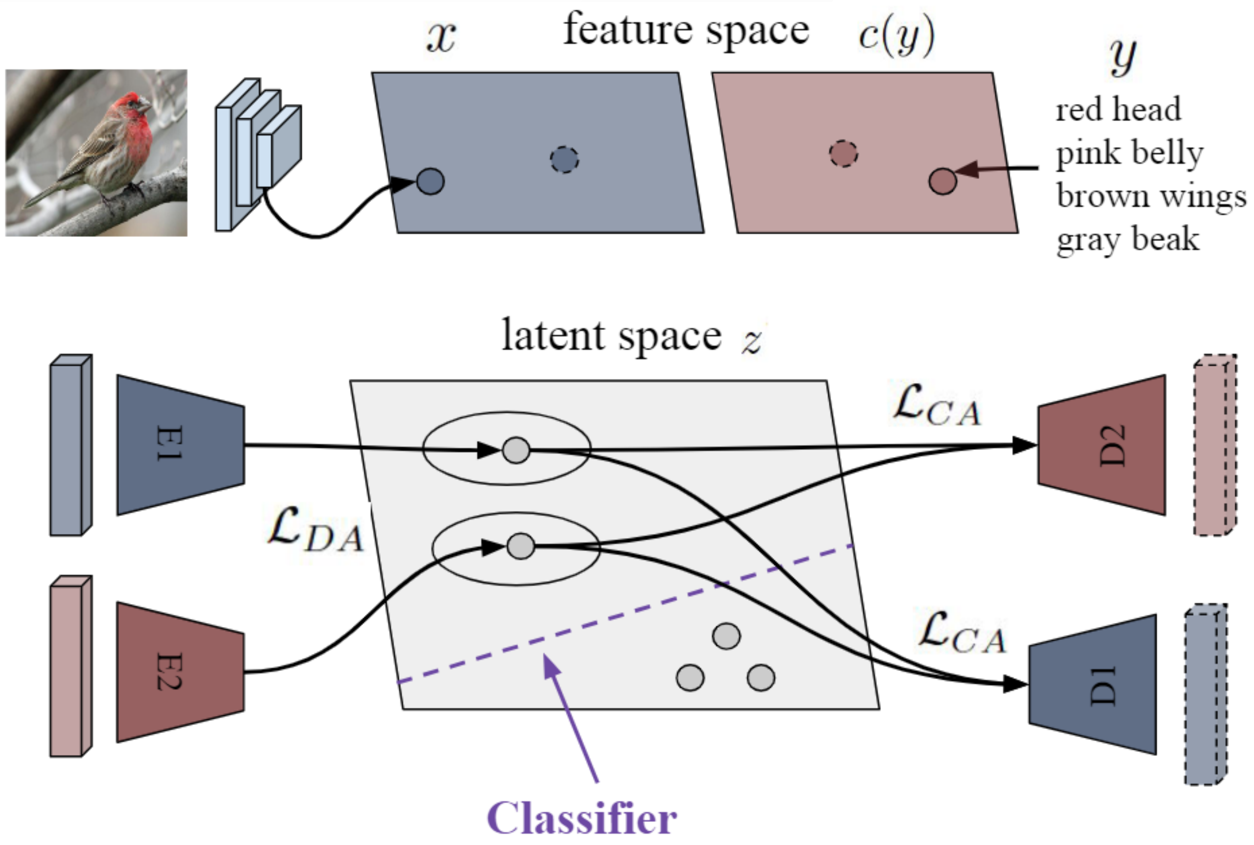
主要思想

1. 训练一个生成模型，该模型能够使用语义信息进行条件生成
2. 向训练好的模型输入 unseen classes 的语义信息，从而生成 unseen 的样本
3. 将训练集中的 seen 样本和生成的 unseen 的样本组合成数据集
4. 将数据集输入分类器进行学习，从而使得分类器能对 seen 和 unseen classes 进行分类



论文思想/贡献

- 通过生成模型构建了视觉和语义统一的隐层空间
- 从统一的隐层空间上采样所有类别的特征，并用于训练分类器
- 做了大量实验，得出了一些重要的结论
- 为基于超球面 OOD 的论文提供了思路



CADA-VAE

VAE Loss

$$\mathcal{L}_{VAE} = \sum_{i}^{M} \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x^{(i)}|z)] - \beta D_{KL}(q_\phi(z|x^{(i)})||p_\theta(z))$$

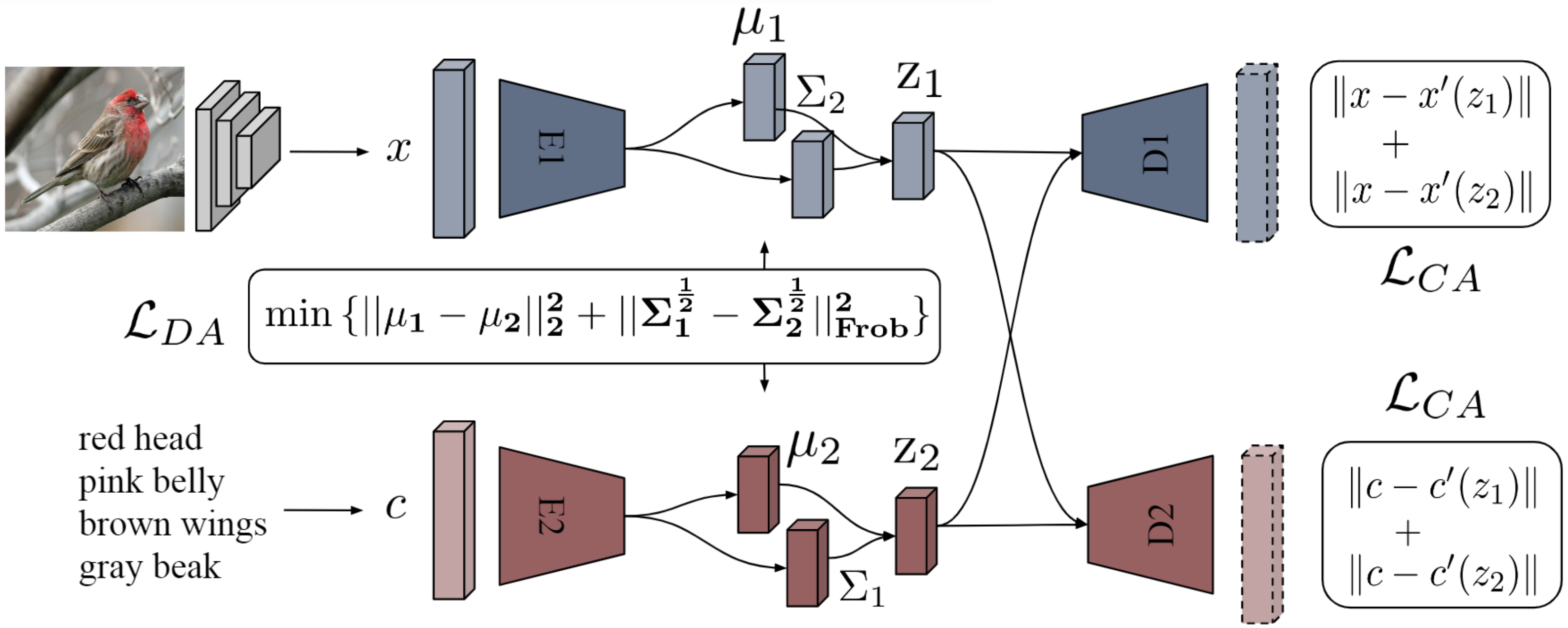


Figure 2: Our Cross- and Distribution Aligned VAE (CADA-VAE). Latent distribution alignment is achieved by minimizing the Wasserstein distance between the latent distributions ($\mathcal{L}_{DA}$). Similarly, the cross-alignment loss ($\mathcal{L}_{CA}$) encourages the latent distributions to align through cross-modal reconstruction.

CADA-VAE

Cross-Alignment Loss

$$\mathcal{L}_{CA} = \sum_{i}^{M}\sum_{j\neq i}^{M}|x^{(j)} - D_j(E_i(x^{(i)}))|$$

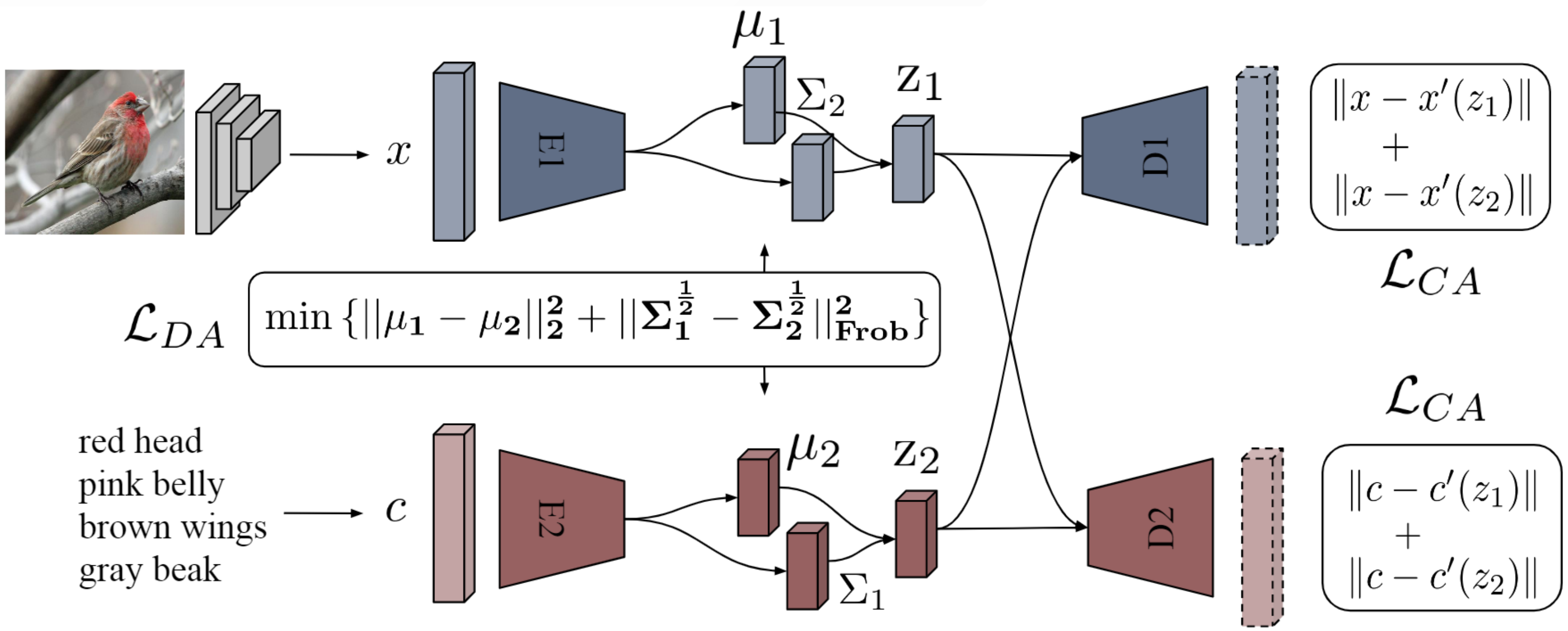


Figure 2: Our Cross- and Distribution Aligned VAE (CADA-VAE). Latent distribution alignment is achieved by minimizing the Wasserstein distance between the latent distributions ($\mathcal{L}_{DA}$). Similarly, the cross-alignment loss ($\mathcal{L}_{CA}$) encourages the latent distributions to align through cross-modal reconstruction.

CADA-VAE

Distribution-Alignment Loss

$$\mathcal{L}_{DA} = \sum_{i}^{M} \sum_{j \neq i}^{M} W_{ij} \qquad W_{ij} = \left(\|\mu_i - \mu_j\|_2^2 + \|\Sigma_i^{\frac{1}{2}} - \Sigma_j^{\frac{1}{2}}\|_{\text{Frobenius}}^2 \right)^{\frac{1}{2}}$$

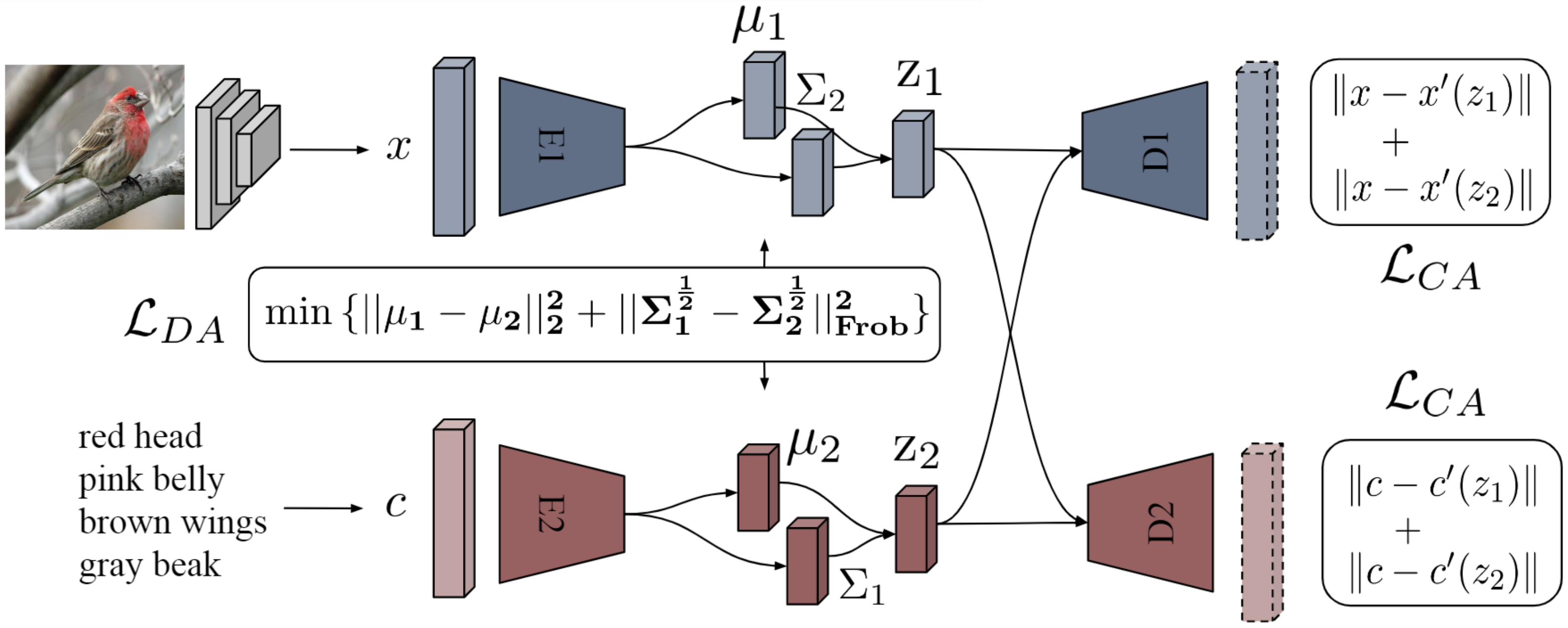


Figure 2: Our Cross- and Distribution Aligned VAE (CADA−VAE). Latent distribution alignment is achieved by minimizing the Wasserstein distance between the latent distributions ($\mathcal{L}_{DA}$). Similarly, the cross-alignment loss ($\mathcal{L}_{CA}$) encourages the latent distributions to align through cross-modal reconstruction.

CADA-VAE

Total Loss

$$\mathcal{L}_{CADA-VAE} = \mathcal{L}_{VAE} + \gamma\mathcal{L}_{CA} + \delta\mathcal{L}_{DA}$$

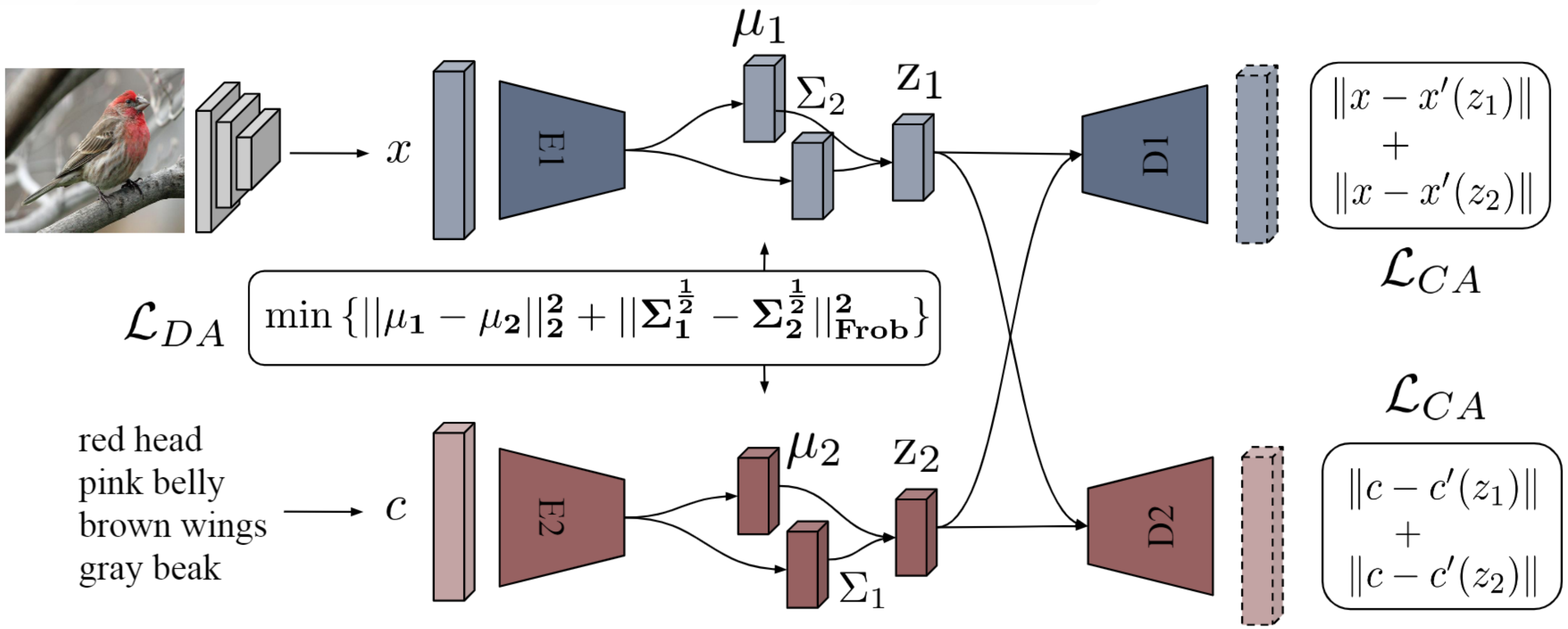


Figure 2: Our Cross- and Distribution Aligned VAE (CADA-VAE). Latent distribution alignment is achieved by minimizing the Wasserstein distance between the latent distributions ($\mathcal{L}_{DA}$). Similarly, the cross-alignment loss ($\mathcal{L}_{CA}$) encourages the latent distributions to align through cross-modal reconstruction.

CADA-VAE

分类

- 训练
  - Seen 样本：$z_{x_s} = E_v(x_s)$
  - Unseen 样本：$z_{a_u} = E_a(a_u)$
  - 分类器：$p(z) \rightarrow y$
- 测试
  - $z = E_v(x) \quad p(z) \rightarrow y$

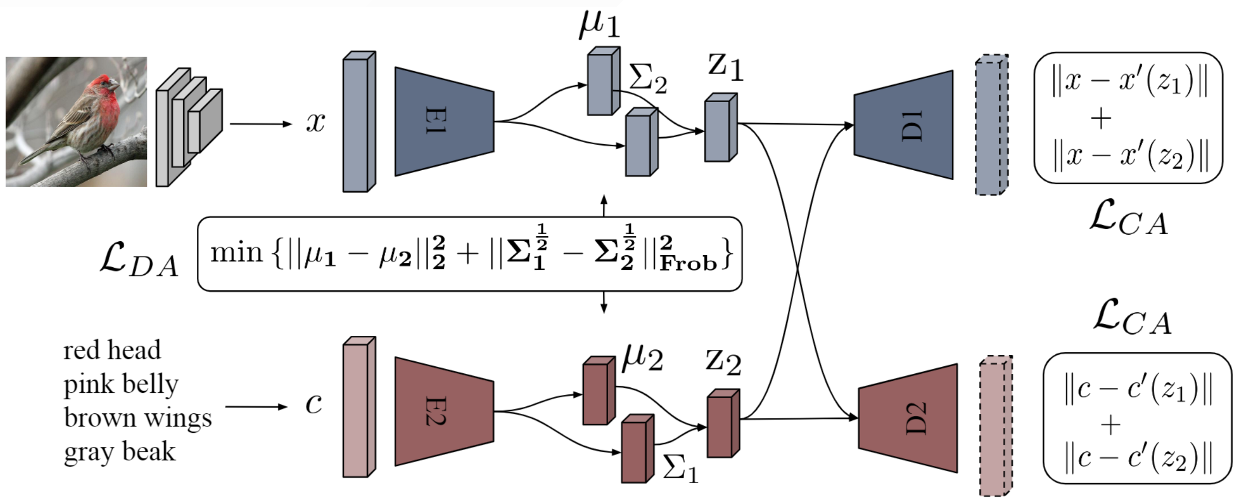


Figure 2: Our Cross- and Distribution Aligned VAE (CADA-VAE). Latent distribution alignment is achieved by minimizing the Wasserstein distance between the latent distributions ($\mathcal{L}_{DA}$). Similarly, the cross-alignment loss ($\mathcal{L}_{CA}$) encourages the latent distributions to align through cross-modal reconstruction.

Experiments

Comparison with SOTA

Model	Feature Size	CUB S	CUB U	CUB H	SUN S	SUN U	SUN H	AWA1 S	AWA1 U	AWA1 H	AWA2 S	AWA2 U	AWA2 H
CMT [27]	2048	49.8	7.2	12.6	21.8	8.1	11.8	87.6	0.9	1.8	90.0	0.5	1.0
SJE [2]		59.2	23.5	33.6	30.5	14.7	19.8	74.6	11.3	19.6	73.9	8.0	14.4
ALE [1]		62.8	23.7	34.4	33.1	21.8	26.3	76.1	16.8	27.5	81.8	14.0	23.9
LATEM [34]		57.3	15.2	24.0	28.8	14.7	19.5	71.7	7.3	13.3	77.3	11.5	20.0
EZSL [24]		63.8	12.6	21.0	27.9	11.0	15.8	75.6	6.6	12.1	77.8	5.9	11.0
SYNC [4]		70.9	11.5	19.8	43.3	7.9	13.4	87.3	8.9	16.2	90.5	10.0	18.0
DeViSE [6]		53.0	23.8	32.8	27.4	16.9	20.9	68.7	13.4	22.4	74.7	17.1	27.8
f-CLSWGAN [36]	1024	57.7	43.7	49.7	36.6	42.6	39.4	61.4	57.9	59.6	68.9	52.1	59.4
CVAE [18]		–	–	34.5	–	–	26.7	–	–	47.2	–	–	51.2
SE [14]		53.3	41.5	46.7	30.5	40.9	34.9	67.8	56.3	61.5	68.1	58.3	62.8
ReViSE [29]	75/100	28.3	37.6	32.3	20.1	24.3	22.0	37.1	46.1	41.1	39.7	46.4	42.8
ours (CADA-VAE)	64	53.5	51.6	**52.4**	35.7	47.2	**40.6**	72.8	57.3	**64.1**	75.0	55.8	**63.9**

Experiments

Ablation Study

Model	S	U	H
DA–VAE	48.1	43.8	45.8
CA–VAE	52.6	48.1	50.2
CADA–VAE	**53.5**	**51.6**	**52.4**

Table 1: Ablation study. We compare GZSL accuracy on CUB for different multi-modal alignment objective functions, i.e. DA–VAE (distribution aligned VAE) , CA–VAE (cross-aligned VAE) and CADA–VAE (cross and distribution aligned VAE).

Experiments

Class Embedding

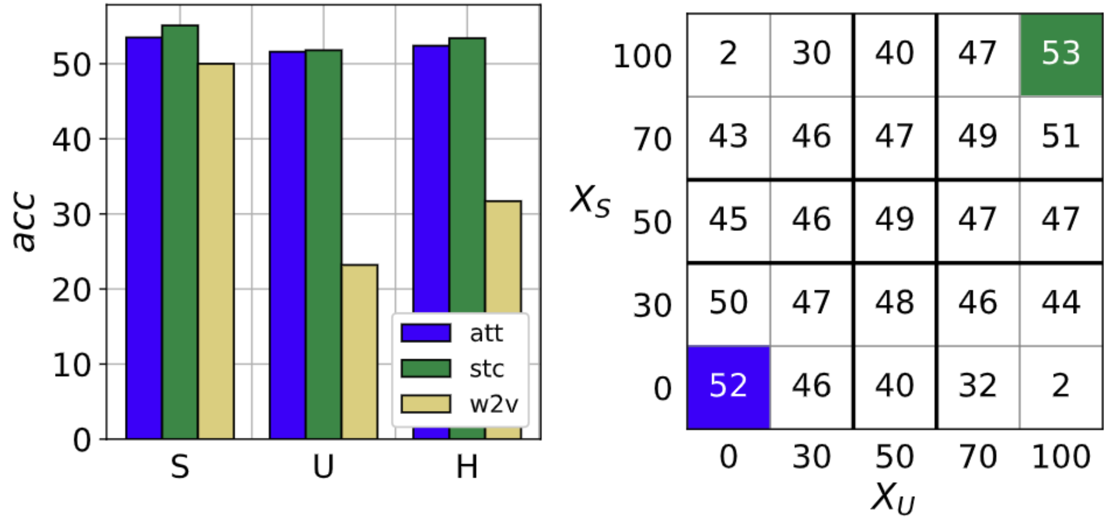


Figure 3: Effect of different class embeddings. (Left) Seen, unseen and harmonic mean accuracy for CUB using different class embeddings as side information. (Right) Using both attributes and sentences as side information, i.e. X_S: the percentage of seen classes with sentences, X_U: the percentage of unseen classes with sentences. Attributes are the class embeddings for the $(100 - X)\%$ of the classes.

- sentence > attribute > w2v

Experiments

Dimentionality of the Latent Features

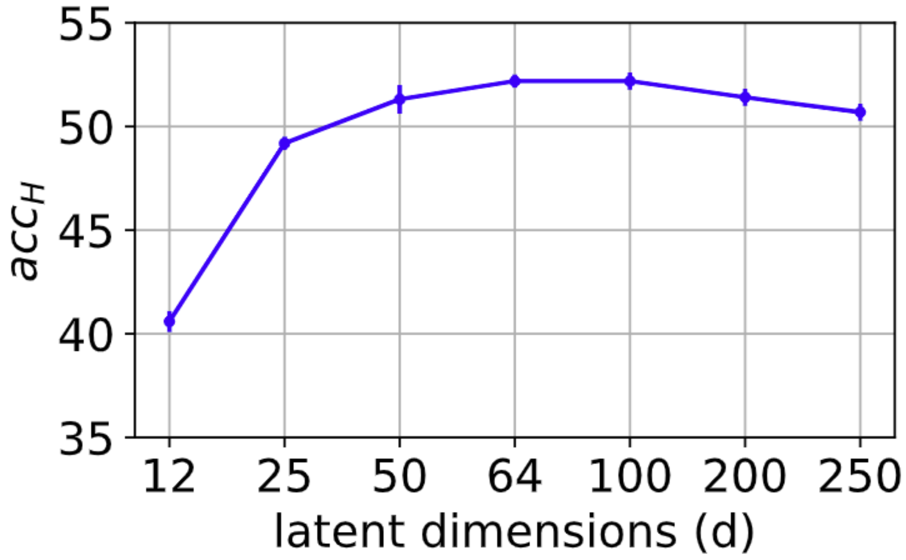


Figure 4: The influence of the dimentionality of the latent features that are generated by CADA−VAE and used to train the GZSL classifier. We measure the harmonic mean accuracy on the CUB dataset

- $d_{\text{best}} = 64$

Experiments

Number of the Latent Features Per Class

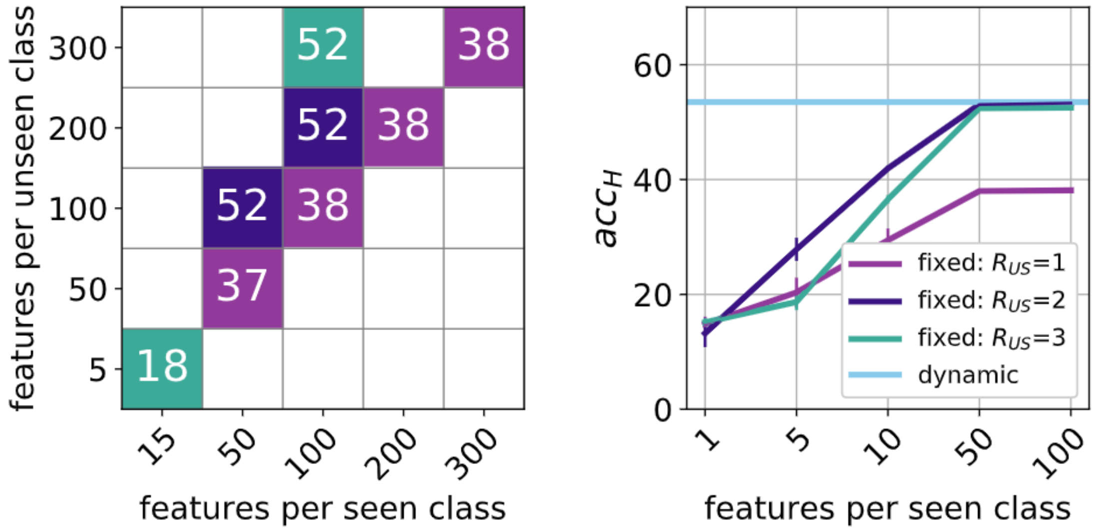


Figure 5: Analyzing the effect of the number of latent features per class on the harmonic mean accuracy in GZSL. An unseen-seen ratio R_{US} of 2 means that twice as many samples are generated for unseen classes than for seen classes . The dynamic dataset (light blue) does not rely on a fixed number of sampled latent features.

- 需要为 unseen 多生成一些样本，在 CUB 中最佳为 $N_S : N_U = 1 : 2$

Experiments

Few Shot Settings

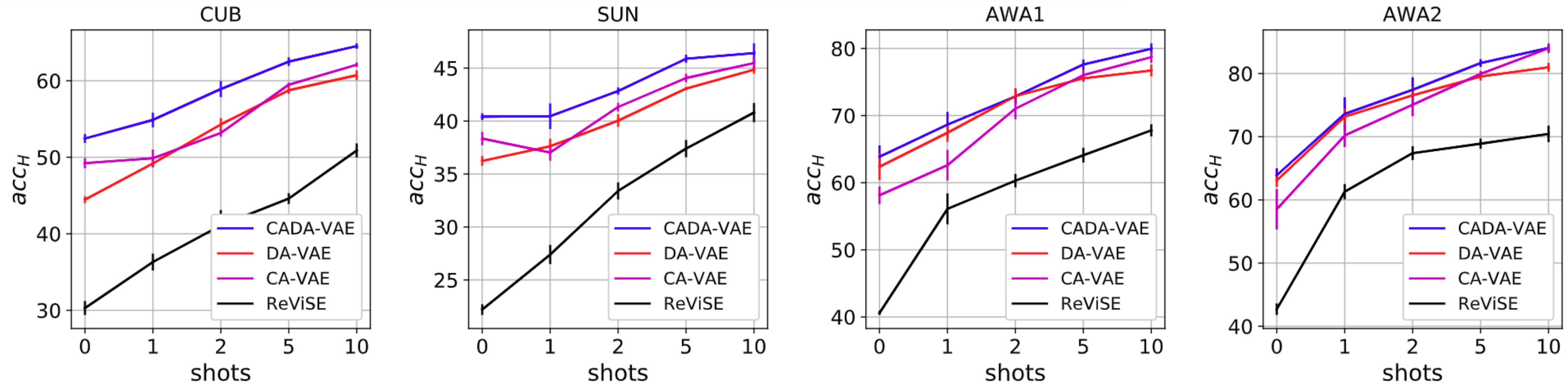


Figure 6: Comparing CA−VAE, DA−VAE, CADA−VAE with ReViSE [30] with increasing numbers of training samples from unseen classes, i.e. in the generalized few-shot setting.

- CUB、SUN 上的提升大于 AWA1、AWA2，说明粗粒度数据集的图像特征已经具有足够的判别力

Out-of-distribution Detection Baseon CADA-VAE

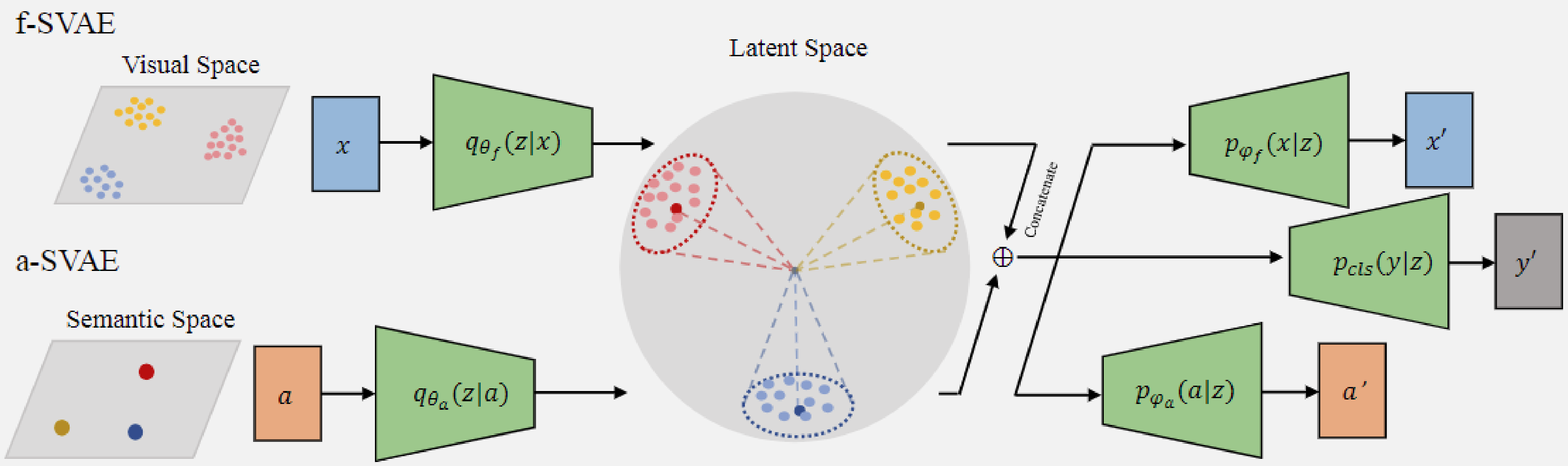


Fig. 2. Our model consists of two SVAEs, one for visual features and another for semantic attributes. By combining the objective functions of the two SVAEs with a cross-reconstruction loss and a classification loss, we train our model to align the latent distributions of visual features and semantic attributes class-wisely. In this way, each class can be represented by a vMF distribution whose boundary is easy to find.