A Boundary Based Out-of-Distribution Classifier for Generalized Zero-Shot Learning

Reporter：陈思玉

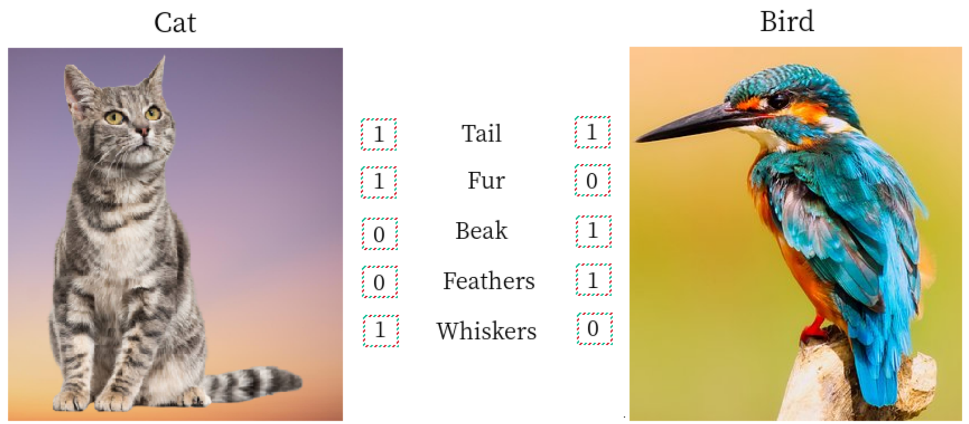
2023.6.10

Chen X, Lan X, Sun F, et al. A boundary based out-of-distribution classifier for generalized zero-shot learning[C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIV 16. Springer International Publishing, 2020: 572-588.

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

OOD-based Methods

主要思想

1. 训练一个区分 seen/unseen 的分类器（**重点**）
2. 对于 seen 样本，训练一个普通的分类器，例如 softmax
3. 对于 unseen 样本，使用特殊的方法
4. 测试阶段，根据 seen/unseen 分类器的分类结果，选择 seen/unseen 专家来分类

难点

- 由于缺乏 unseen 样本，将未知的 unseen 类别样本与已有的 seen 类别样本区分开是一个挑战
- 假设能够完全分开，对于 unseen 样本该如何分类

A Boundary Based Out-of-Distribution Classifier for Generalized Zero-Shot Learning

Chen X, Lan X, Sun F, et al. A boundary based out-of-distribution classifier for generalized zero-shot learning[C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIV 16. Springer International Publishing, 2020: 572-588.

Abstract

1. seen/unseen 分类器：在单位超球体上学习一个共享的潜在空间，其中视觉特征和语义属性的潜在分布按类别对齐，然后找到每个类的流形的边界和中心，通过利用类中心和边界，可以将 unseen 的样本与 seen 的样本分开
2. seen 专家：encoder + softmax
3. unseen 专家：其他 GZSL 方法的 unseen 分类器

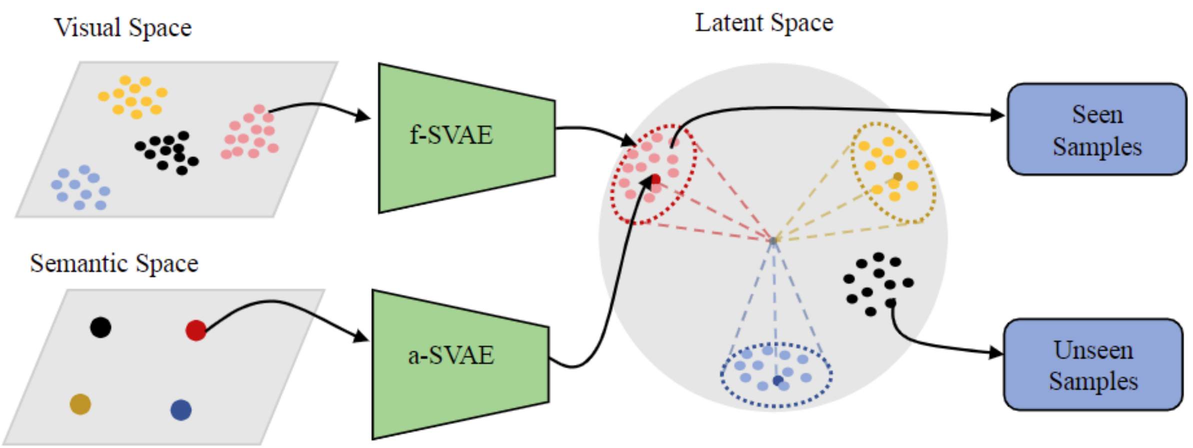


Fig. 1. The boundary based OOD classifier learns a bounded manifold for each seen class on a unit hyper-sphere (latent space). By using the manifold boundaries (dotted circles) and the centers (dark-colored dots), the unseen samples (black dots) can be separated from the seen samples (colored dots).

VAE

ELBO

$$\log p_\phi(\mathbf{x}) - D_{\mathrm{KL}}(q_\theta(\mathbf{z}|\mathbf{x})\|p_\phi(\mathbf{z}|\mathbf{x})) = \mathbb{E}_{\mathbf{z}\sim q_\theta(\mathbf{z}|\mathbf{x})}[\log p_\phi(\mathbf{x}|\mathbf{z})] - D_{\mathrm{KL}}(q_\theta(\mathbf{z}|\mathbf{x})\|p_\phi(\mathbf{z}))$$

- 先验：$p_\theta(\mathbf{z})$
- 似然：$p_\theta(\mathbf{x}|\mathbf{z})$
- 后验：$p_\theta(\mathbf{z}|\mathbf{x})$

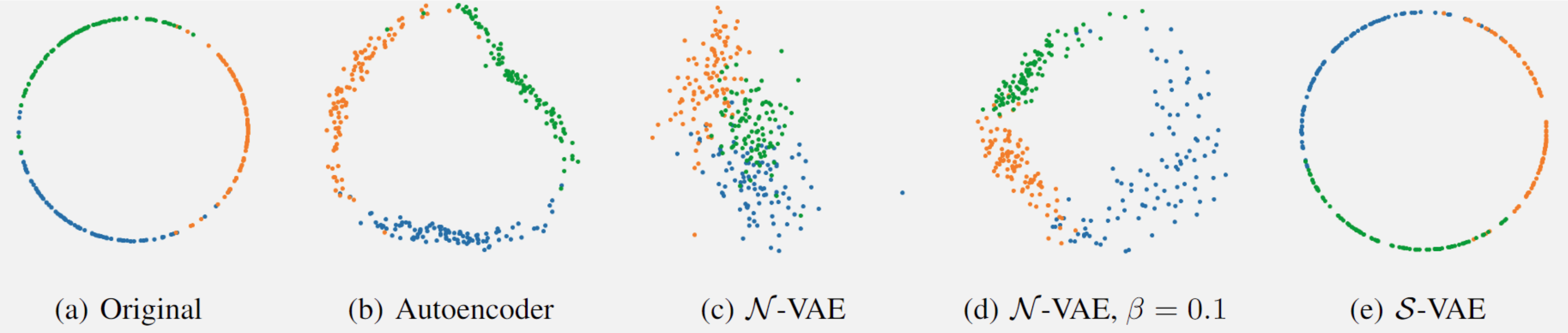


Figure 1: Plots of the original latent space (a) and learned latent space representations in different settings, where β is a re-scaling factor for weighting the KL divergence. (Best viewed in color)

分布

其中，先验一般选择**正态分布** $\mathcal{N} \sim (\mu, \sigma^2)$，然后让后验去接近先验

但是正态分布倾向于将分布拉向原点，对于原本分布为类似球形的数据分布的重建效果不太好

S-VAE

von Mises-Fisher (vMF) 分布

$$q(\mathbf{z}|\mu,\kappa) = \mathcal{C}_m(\kappa)\exp(\kappa\mu^T\mathbf{z})$$

$$\mathcal{C}_m(\kappa) = \frac{\kappa^{m/2-1}}{(2\pi)^{m/2}\mathcal{I}_{m/2-1}(\kappa)}$$

- $\|\mu\|^2 = 1$ 表示球面上的方向
- $\kappa \in \mathbb{R}_{\geq 0}$ 表示 μ 周围的浓度
- $\mathcal{C}_m(\kappa)$ 是归一化常数
- $\mathcal{I}_v$ 表示 v 阶第一类修正贝塞尔函数

Boundary Based Out-of-Distribution Classifier

- 为视觉特征和语义属性构建潜在空间，并进行对齐
- seen 类别学习一个有界流形以及其边界，每个类别可以用 1 个 vMF 分布表示
- 通过边界，可以确定测试样本是否被投影到流形中，若可以则判定为 seen，否则为 unseen

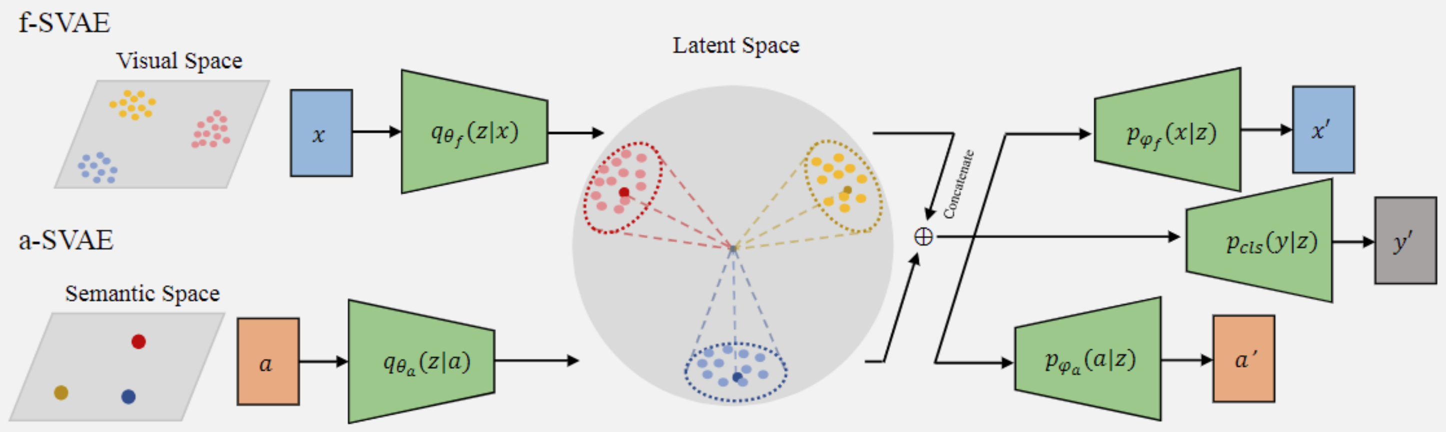


Fig. 2. Our model consists of two SVAEs, one for visual features and another for semantic attributes. By combining the objective functions of the two SVAEs with a cross-reconstruction loss and a classification loss, we train our model to align the latent distributions of visual features and semantic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

Boundary Based Out-of-Distribution Classifier

对齐视觉、语义

$$L_{f-SVAE} = \mathbb{E}_{p(x,a)}[\mathbb{E}_{q_{\theta_f}(z|x)}[\log p_{\phi_f}(x|z)] - \lambda_f D_z(q_{\theta_f}(z|x)\|q_{\theta_a}(z|a))]$$

$$L_{a-SVAE} = \mathbb{E}_{p(x,a)}[\mathbb{E}_{q_{\theta_a}(z|a)}[\log p_{\theta_a}(a|z)] - \lambda_a D_z(q_{\theta_a}(z|a)\|q_{\theta_f}(z|x))]$$

- 其中 $D_z(q_{\theta_f}(z|x)\|q_{\theta_a}(z|a))$ 是 2 个分布之间的 Earth Mover’ s Distance (EMD)
  - 不用 KL 散度的原因是因为当 2 个区域部分重合时效果不好

$$D_z(q_{\theta_f}(z|x)\|q_{\theta_a}(z|a)) = \inf_{\Omega\in\prod(q_{\theta_f},q_{\theta_a})} \mathbb{E}_{(z_1,z_2)\sim\Omega}[\|z_1 - z_2\|]$$

交叉重建

$$L_{cr} = \mathbb{E}_{p(x,a)}[\mathbb{E}_{q_{\theta_a}(z|a)}[\log p_{\phi_f}(x|z)] + \mathbb{E}_{q_{\theta_f}(z|x)}[\log p_{\phi_a}(a|z)]]$$

分类损失

$$L_{cls} = \mathbb{E}_{p(x,y,a)}[\mathbb{E}_{q_{\theta_a}(z|a)}[\log p_{\phi_{cls}}(y|z)] + \mathbb{E}_{q_{\theta_f}(z|x)}[\log p_{\phi_{cls}}(y|z)]]$$

Boundary

类中心

对于类 $y^i \in \mathcal{Y}_s$，可以使用其语义信息找到类中心

- 给定 $a^i \in \mathcal{A}_s$，a-SVAE 预测一个 vMF 分布 $q(z|\mu(a^i), \kappa(a^i))$，其中 $\mu(a^i)$ 被视为类中心。

类边界

对于类边界，

- 将所有可见类的训练样本编码为潜在变量 z^i
- 计算每个潜在变量 z^i 和相应的类中心 $\mu(a^i)$ 之间的余弦相似度 $S(z^i, \mu(a^i))$
- 搜索 1 个阈值 η 来表示边界

$$y^{OOD} = \begin{cases} \text{unseen}, & \text{if} \quad \max\{S(z, \mu(a^i)) | \forall a^i \in \mathcal{A}_s\} < \eta \\ \text{seen}, & \text{if} \quad \max\{S(z, \mu(a^i)) | \forall a^i \in \mathcal{A}_s\} \geq \eta \end{cases}$$

Experiments

中南大學
CENTRAL SOUTH UNIVERSITY

Table 1. Comparison with various gating models on validation set. **AUC** denotes Area-Under-Curve when sweeping over detection threshold. **FPR** denotes False-Positive-Rate on the threshold that yields 95% True Positive Rate for detecting in-distribution samples. The best results are highlighted with bold numbers.

	AWA1			CUB			SUN		
Method	**H**	**AUC**	**FPR**	**H**	**AUC**	**FPR**	**H**	**AUC**	**FPR**
MAX-SOFTMAX-3 [12]	53.1	88.6	56.8	43.6	73.4	79.6	38.4	61.0	92.3
CB-GATING-3[4]	56.8	92.5	45.5	44.8	82.0	72.0	40.1	77.7	77.5
Ours	**70.1**	**95.0**	**12.5**	**67.7**	**99.4**	**2.5**	**71.0**	**99.5**	**1.6**

Table 2. OOD classification results of our approach by selecting different thresholds using γ.

	AWA1		AWA2		CUB		FLO		SUN	
	TPR	**FPR**	**TPR**	**FPR**	**TPR**	**FPR**	**TPR**	**FPR**	**TPR**	**FPR**
$\gamma = 0.85$	85.0	5.3	85.2	6.8	84.2	0.7	85.3	0.4	85.4	0.2
$\gamma = 0.90$	90.1	6.3	89.8	8.2	89.5	0.9	88.2	0.6	90.6	0.2
$\gamma = 0.95$	95.4	7.9	95.2	10.6	94.9	1.1	94.4	0.8	95.1	0.4

Experiments

中南大學
CENTRAL SOUTH UNIVERSITY

Table 3. Generalized Zero-Shot Learning results on AWA1, AWA2, CUB, FLO and SUN. We measure the AP of Top-1 accuracy in %. The best results are highlighted with bold numbers.

	AWA1			AWA2			CUB			FLO			SUN		
Method	**ts**	**tr**	**H**	**ts**	**tr**	**H**	**ts**	**tr**	**H**	**ts**	**tr**	**H**	**ts**	**tr**	**H**
SJE [2]	11.3	74.6	19.6	8.0	73.9	14.4	23.5	59.2	33.6	13.9	47.6	21.5	14.7	30.5	19.8
ALE [1]	16.8	76.1	27.5	14.0	81.8	23.9	23.7	62.8	34.4	13.3	61.6	21.9	21.8	33.1	26.3
PSR [3]	-	-	-	20.7	73.8	32.3	24.6	54.3	33.9	-	-	-	20.8	37.2	26.7
SAE [14]	16.7	82.5	27.8	8.0	73.9	14.4	18.8	58.5	29.0	-	-	-	8.8	18.0	11.8
ESZSL [25]	6.6	75.6	12.1	5.9	77.8	11.0	12.6	63.8	21.0	11.4	56.8	19.0	11.0	27.9	15.8
LESAE [18]	19.1	70.2	30.0	21.8	70.6	33.3	24.3	53.0	33.3	-	-	-	21.9	34.7	26.9
ReViSE [28]	46.1	37.1	41.1	46.4	39.7	42.8	37.6	28.3	32.3	-	-	-	24.3	20.1	22.0
CMT [27]	0.9	87.6	1.8	0.5	90.0	1.0	7.2	49.8	12.6	-	-	-	8.1	21.8	11.8
SYNC [5]	8.9	87.3	16.2	10.0	90.5	18.0	11.5	70.9	19.8	-	-	-	7.9	43.3	13.4
DeViSE [10]	13.4	68.7	22.4	17.1	74.7	27.8	23.8	53.0	32.8	9.9	44.2	16.2	16.9	27.4	20.9
CRnet [33]	58.1	74.7	65.4	52.6	78.8	63.1	45.5	56.8	50.5	-	-	-	34.1	36.5	35.3
CVAE [21]	-	-	47.2	-	-	51.2	-	-	34.5	-	-	-	-	-	26.7
SP-AEN [6]	-	-	-	23.3	90.9	37.1	34.7	70.6	46.6	-	-	-	24.9	38.6	30.3
f-CLSWGAN [31]	57.9	61.4	59.6	52.1	68.9	59.4	43.7	57.7	49.7	59.0	73.8	65.6	42.6	36.6	39.4
cycle-(U)WGAN [9]	59.6	63.4	59.8	-	-	-	47.9	59.3	53.0	61.6	69.2	65.2	47.2	33.8	39.4
SE [15]	56.3	67.8	61.5	58.3	68.1	62.8	41.5	53.3	46.7	-	-	-	40.9	30.5	34.9
CADA-VAE [26]	57.3	72.8	64.1	55.8	75.0	63.9	51.6	53.5	52.4	-	-	-	47.2	35.7	40.6
AFC-GAN [17]	-	-	-	58.2	66.8	62.2	53.5	59.7	56.4	60.2	80.0	68.7	49.1	36.1	41.6
COSMO+fCLSWGAN [4]	64.8	51.7	57.5	-	-	-	41.0	60.5	48.9	59.6	81.4	68.8	35.3	40.2	37.6
COSMO+LAGO [4]	52.8	80.0	63.6	-	-	-	44.4	57.8	50.2	-	-	-	44.9	37.7	41.0
Ours($\gamma = 0.95$)	59.0	94.3	**72.6**	55.9	94.9	**70.3**	53.8	94.6	**68.6**	61.9	91.7	**73.9**	57.8	95.1	**71.9**

Experiments

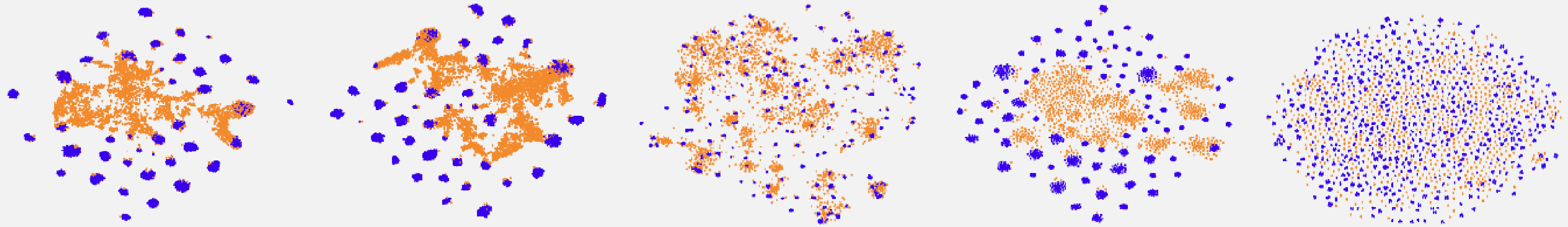


Fig. 4. The t-SNE visualization results for the learned latent space on the test sets of AWA1, AWA2, CUB, FLO and SUN. The blue dots represent the variables encoded from seen classes. The orange dots represent the variables encoded from unseen classes.

Table 4. Binary classification results of different training objective functions. We report the AUC and the FPR corresponding to $\gamma = 0.95$.

	AWA1		CUB	
Objective Function	**AUC**	**FPR**	**AUC**	**FPR**
$L_{f-SVAE} + L_{a-SVAE}$	62.5	93.3	56.1	88.5
$L_{f-SVAE} + L_{a-SVAE} + L_{cr}$	89.3	44.2	60.6	86.7
$L_{f-SVAE} + L_{a-SVAE} + L_{cls}$	94.9	15.7	98.2	9.2
$L_{overall}$	96.8	7.9	99.6	1.1