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

Reporter：陈思玉

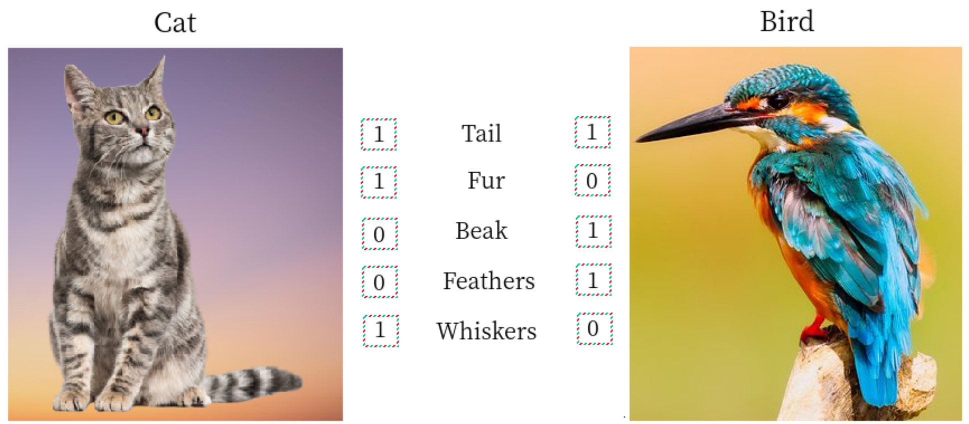
2023.6.10

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

Adaptive Confidence Smoothing for Generalized Zero-Shot Learning

Atzmon Y, Chechik G. Adaptive confidence smoothing for generalized zero-shot learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 11671-11680.

Abstract

中南大學 CENTRAL SOUTH UNIVERSITY

1. seen/unseen 分类器：使用 1 个门控模型来输出样本属于 seen 和 unseen 的概率，并将其作为权重并分别赋予 seen 专家和 unseen 专家
2. seen 专家：softmax
3. unseen 专家：其他 GZSL 方法的 unseen 分类器

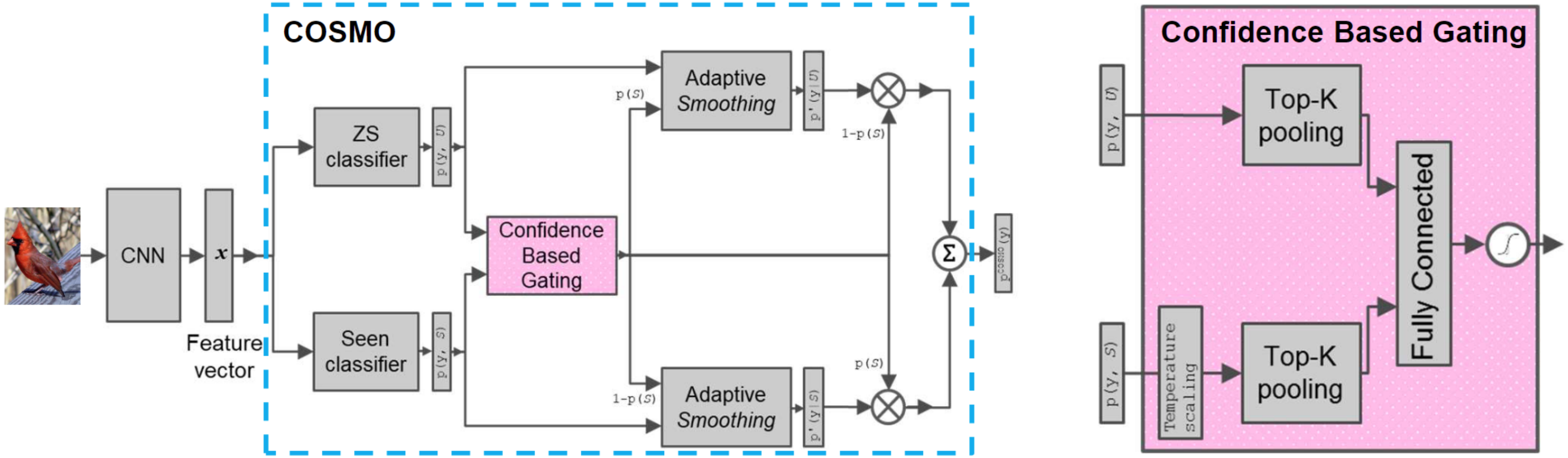


Figure 2. **Left**, COSMO Architecture: We decompose the GZSL task into three sub-tasks that can be addressed separately. (1) A model trained to classify seen $\mathcal{S}$ classes. (2) A model classifying *unseen* $\mathcal{U}$ classes, namely a ZSL model, conditioned on $\mathcal{U}$. (3) A *gating* binary classifier trained to discriminate between seen and unseen classes and to weigh the two models in a soft way; Before weighing (1) & (2) softmax distributions, we add a prior for each if the gating network provides low confidence (Figure 1 and Sec 4.2). **Right**, The gating network (Zoom-in): It takes softmax scores as inputs. We train it to be aware of the response of softmax scores to *unseen* images, with samples from held-out classes. Because test classes are different from train classes, we pool the top-K scores, achieving invariance to class identity (Section 4.1). The fully-connected layer only learns 10-50 weights (K is small) since this is a binary classifier.

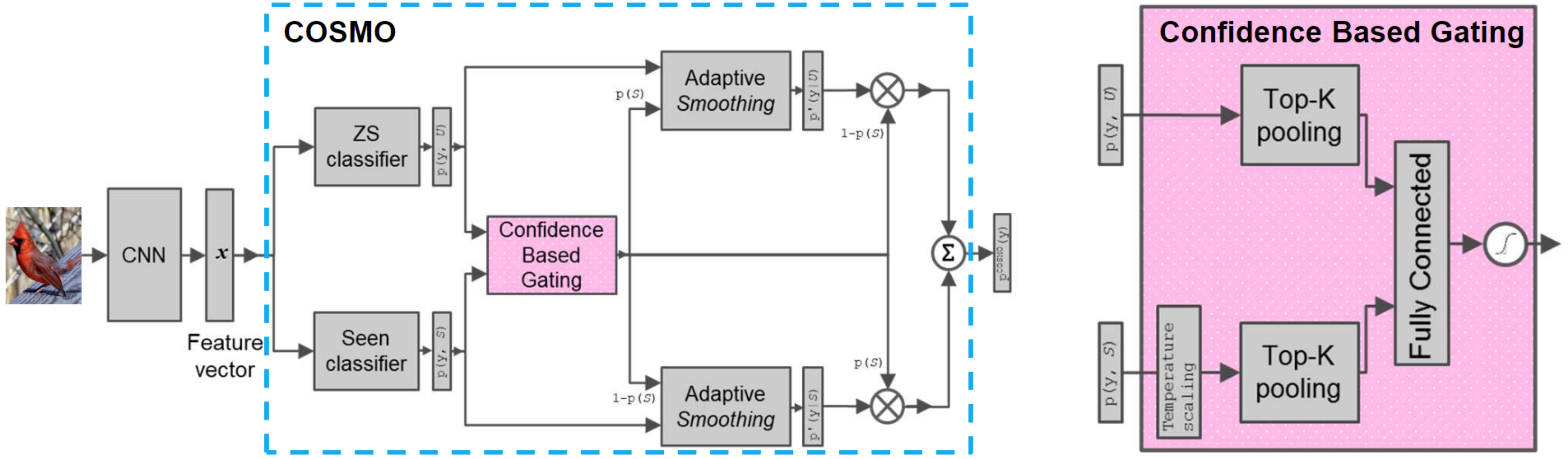
Confidence Based Gating

结构

- 输入：seen 专家和 unseen 专家的输出
- 各自选取 Top-K 个输出，并拼接
- 输出：seen 和 unseen 的概率

训练

由于缺乏 unseen 图像，该论文将从未用于训练的 seen 类中创建了一个保留类 $\mathcal{H}$，并使用它们来估计对 unseen 图像的输出响应



Adaptive Confidence Smoothing

由于希望对于专家（例如 seen 专家）范围外的样本（例如 unseen），该专家能够输出所有类别均偏低的概率，该论文将专家预测置信度与先验分布相结合

$$p^{\lambda}(y|\mathcal{U}) = (1-\lambda)p(y|\mathcal{U}) + \lambda\pi^{\mathcal{U}}$$

$$p^{\lambda}(y|\mathcal{S}) = (1-\lambda)p(y|\mathcal{S}) + \lambda\pi^{\mathcal{S}}$$

其中 π 为不以 $\mathbf{x}$ 为条件的先验分布。该论文将先验设置为最大熵分布，即均匀分布

$$\pi^{\mathcal{U}} = 1/(\#\text{unseen classes})$$

$$\pi^{\mathcal{S}} = 1/(\#\text{seen classes})$$

对于 λ 的设置，该论文使用门控模型的输出来作为 seen 和 unseen 的 λ 值

Experiments

中南大學 CENTRAL SOUTH UNIVERSITY

DATASET	AWA Acc_{ts}	AWA Acc_{tr}	AWA Acc_H	SUN Acc_{ts}	SUN Acc_{tr}	SUN Acc_H	CUB Acc_{ts}	CUB Acc_{tr}	CUB Acc_H	FLOWER Acc_{ts}	FLOWER Acc_{tr}	FLOWER Acc_H
NON-GENERATIVE MODELS												
ESZSL [36]	6.6	75.6	12.1	11	27.9	15.8	12.6	63.8	21	11.4	56.8	19
SJE [2]	11.3	74.6	19.6	14.7	30.5	19.8	23.5	59.2	33.6	13.9	47.6	21.5
DEVISE [12]	13.4	68.7	22.4	16.9	27.4	20.9	23.8	53	32.8	9.9	44.2	16.2
SYNC [8]	8.9	87.3	16.2	7.9	43.3	13.4	11.5	70.9	19.8	-	-	-
ALE [1]	16.8	76.1	27.5	21.8	33.1	26.3	23.7	62.8	34.4	34.4	13.3	21.9
DEM [52]	32.8	84.7	47.3	-	-	-	19.6	57.9	29.2	-	-	-
KERNEL [50]	18.3	79.3	29.8	19.8	29.1	23.6	19.9	52.5	28.9	-	-	-
ICINESS [14]	-	-	-	-	-	30.3	-	-	41.8	-	-	-
TRIPLE [51]	27	67.9	38.6	22.2	38.3	28.1	26.5	62.3	37.2	-	-	-
RN [41]	31.4	91.3	46.7	-	-	-	38.1	61.1	47	-	-	-
GENERATIVE MODELS												
SE-GZSL [5]	56.3	67.8	61.5	40.9	30.5	34.9	41.5	53.3	46.7	-	-	-
FCLSWGAN [46]	59.7	61.4	59.6	42.6	36.6	39.4	43.7	57.7	49.7	59	73.8	65.6
FCLSWGAN* (BY PROVIDED CODE)	53.6	67	59.6	40.1	36	37.9	45.1	55.5	49.8	58.1	73.2	64.8
CYCLE-(U)WGAN [10]	59.6	63.4	59.8	47.2	33.8	39.4	47.9	59.3	**53.0**	61.6	69.2	65.2
COSMO AND BASELINES												
CMT [39]	8.4	86.9	15.3	8.7	28	13.3	4.7	60.1	8.7	-	-	-
DCN [27]	25.5	84.2	39.1	25.5	37	30.2	28.4	60.7	38.7	-	-	-
LAGO [7]	21.8	73.6	33.7	18.8	33.1	23.9	24.6	64.8	35.6	-	-	-
CS [9] + LAGO	45.4	68.2	54.5	41.7	25.9	31.9	43.1	53.7	47.9	-	-	-
OURS: COSMO+FCLSWGAN*	64.8	51.7	57.5	35.3	40.2	37.6	41.0	60.5	48.9	59.6	81.4	**68.8**
OURS: COSMO+LAGO	52.8	80	**63.6**	44.9	37.7	**41.0**	44.4	57.8	**50.2**	-	-	-

Table 1. Comparing COSMO with state-of-the-art GZSL non-generative models and with generative models that synthesize feature vectors. Acc_{tr} is the accuracy of seen classes, Acc_{ts} is the accuracy of unseen classes and Acc_H is their harmonic mean. COSMO+LAGO uses LAGO [7] as a baseline GZSL model, and respectively COSMO+fCLSWGAN uses fCLSWGAN [46]. COSMO+LAGO improves Acc_H over state-of-the-art models by 34%, 35%, 7% respectively for AWA, SUN and CUB. Comparing with generative models, COSMO+LAGO closes the non-generative:generative performance gap, and is comparable to or better than these models, while is very easy to train.

Domain-aware Visual Bias Eliminating for Generalized Zero-Shot Learning

Min S, Yao H, Xie H, et al. Domain-aware visual bias eliminating for generalized zero-shot learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 12664-12673.

Abstract

1. seen/unseen 分类器：通过构建互补的语义无关表征和语义相关表征，并依据语义无关表征的熵来判断样本是否属于 seen
2. seen 专家：softmax
3. unseen 专家：通过自动搜索来寻找最佳的语义-视觉对齐架构

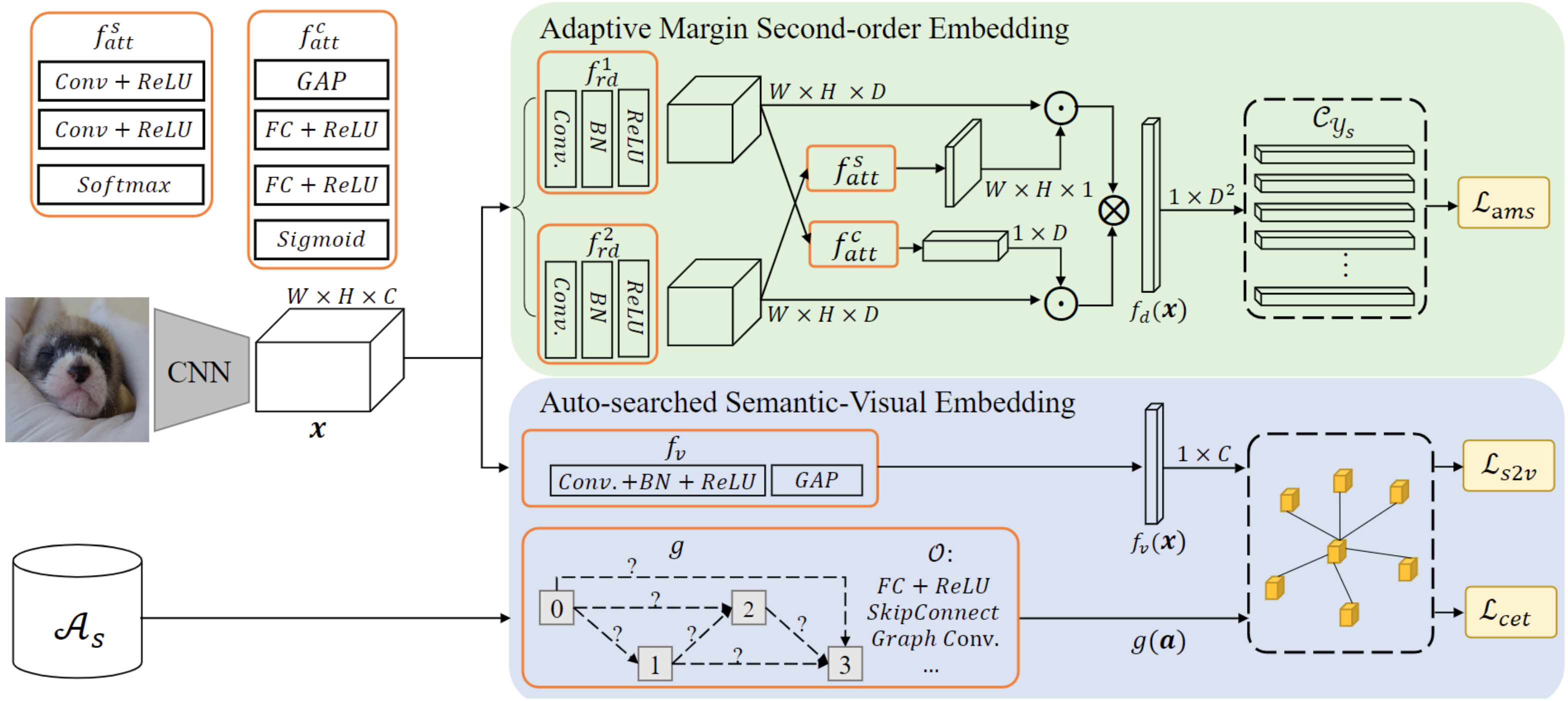


Figure 2. The training framework of DVBE with detailed implementation. GAP is global average pooling, and DVBE is trained with only seen domain data.

DBVE

在传统的语义对齐视觉的基础上额外引入了语义无关表征 $f_d(x)$，并依据其熵来判断 seen/unseen

$$\hat{y} = \arg\min_{y \in \mathcal{Y}_s \cup \mathcal{Y}_u} d(f_v(x), g(a_y))$$

转化成了

$$\hat{y} = \begin{cases} \arg\max_{y \in \mathcal{Y}_s} \mathcal{C}_y(f_d(x)) & \text{if} \quad \mathcal{H}(\mathcal{C}(f_d(x))) \leq \tau \\ \arg\min_{y \in \mathcal{Y}_u} d(f_v(x), g(a(y))) & \text{else} \end{cases}$$

- $\mathcal{C}$： $|\mathcal{Y}_s|$-way 分类器
- $\mathcal{H}(\cdot)$：测量 $f_d(x)$ 的预测分数的熵

$f_d(x)$

该论文没有使用传统的 2048 维视觉特征，而是完整地从 resnet101 网络输入图片进行训练，并将输出的 $W \times H \times C$ 的特征输入网络。

该论文设计了交叉注意力通道来计算 $f_d(x)$

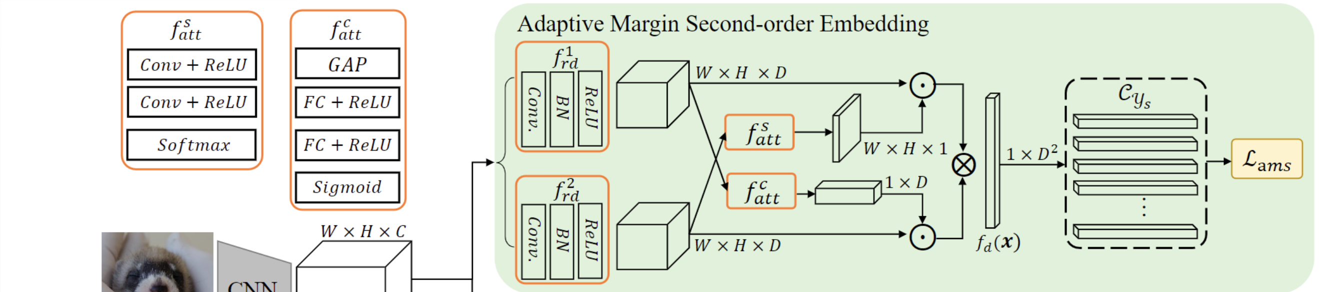
$$f_d(x) = [f^s_{att}(x_2) \odot x_1] \otimes [f^c_{att}(x_1) \odot x_2]$$

$f^s_{att}(\cdot)$ 和 $f^c_{att}(\cdot)$ 分别用于生成空间和通道注意力图 $W \times H \times 1$ 和 $1 \times D$

为了进一步增加 $f_d(x)$ 的区分度，该论文应用自适应边缘 Softmax 来最大化类间差异：

$$\mathcal{L}_{ams} = -\sum_{x \in \mathcal{X}_s} \log \frac{e^{\lambda W_y f_d(x)}}{e^{\lambda W_y f_d(x)} + \sum_{j \in \mathcal{Y}_s, j \neq y} e^{W_j f_d(x)}} \qquad \lambda = e^{-(p_y(x)-1)^2/\sigma^2}$$

其中 W 是分类器权重，y 是真实标签，而 λ 是根据样本难度自适应计算的：



unseen 分类

$$\mathcal{L}_{s2v} = \sum_{x \in \mathcal{X}_s} d(f_v(x), g(a_y))$$

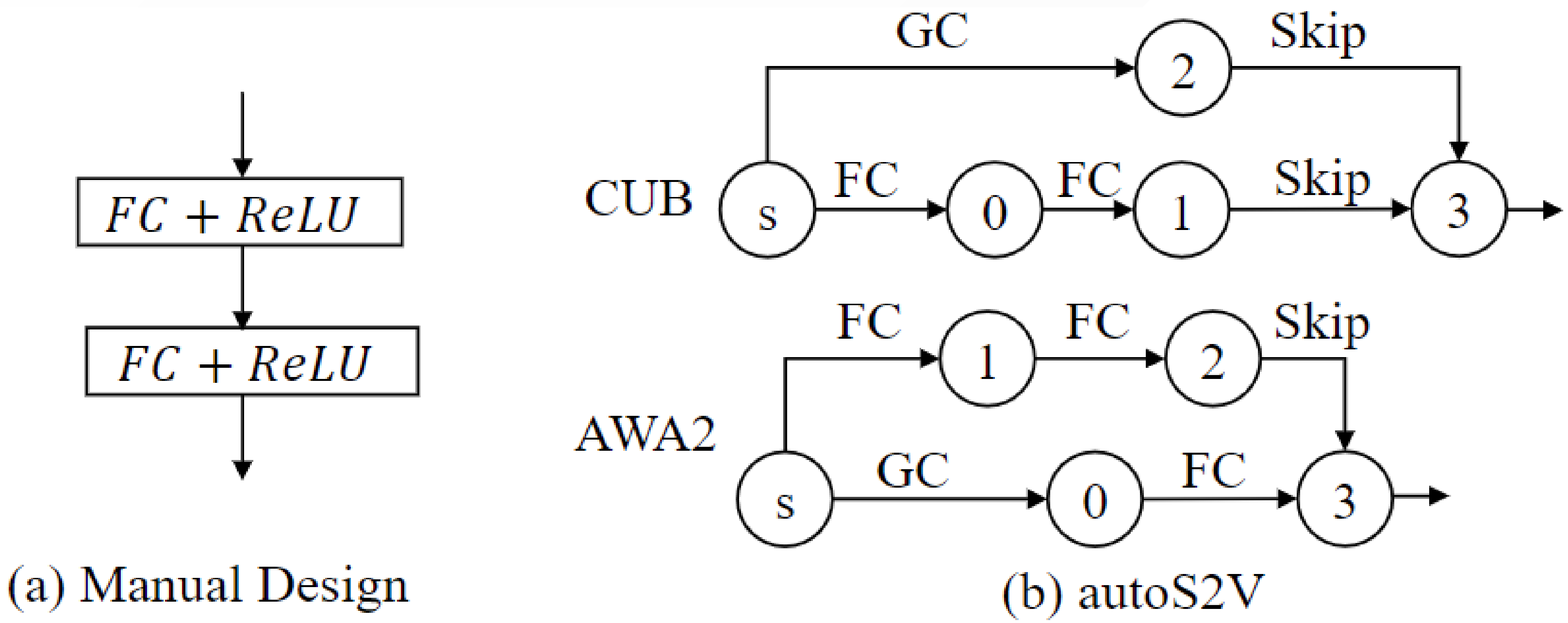


Figure 4. (a) The hand-designed architecture. b) The learned task-specific architectures from autoS2V for CUB and AWA2, respectively. GC is the graph convolution.

Experiments

Table 2. Results of GZSL on four classification benchmarks. Generative methods (GEN) utilizes extra synthetic unseen domain data for training. Since many previous methods cannot be end-to-end trained, we define DVBE and DVBE* as fixing and finetuning the backbone weights, respectively. † indicates the prediction ensemble from global and local regions.

	Methods	CUB [44]			AWA2 [46]			aPY [12]			SUN [36]		
		MCA_u	MCA_s	H	MCA_u	MCA_s	H	MCA_u	MCA_s	H	MCA_u	MCA_s	H
GEN	FGN[47]	43.7	57.7	49.7	-	-	-	-	-	-	42.6	36.6	39.4
	SABR-I[37]	55.0	58.7	56.8	30.3	93.9	46.9	-	-	-	50.7	35.1	41.5
	f-VAEGAN-D2[49]	63.2	75.6	68.9	-	-	-	-	-	-	50.1	37.8	43.1
NON-GEN	CDL[18]	23.5	55.2	32.9	-	-	-	19.8	48.6	28.1	21.5	34.7	26.5
	PSR-ZSL[1]	24.6	54.3	33.9	20.7	73.8	32.2	13.5	51.4	21.4	20.8	37.2	26.7
	SP-AEN[5]	34.7	70.6	46.6	23.3	90.9	37.1	13.7	63.4	22.6	24.9	38.6	30.3
	DLFZRL[41]	-	-	37.1	-	-	45.1	-	-	31.0	-	-	24.6
	MLSE[9]	22.3	71.6	34.0	23.8	83.2	37.0	12.7	74.3	21.7	20.7	36.4	26.4
	TripletLoss[4]	55.8	52.3	53.0	48.5	83.2	61.3	-	-	-	47.9	30.4	36.8
	COSMO[2]	44.4	57.8	50.2	-	-	-	-	-	-	44.9	37.7	41.0
	PREN*[55]	32.5	55.8	43.1	32.4	88.6	47.4	-	-	-	35.4	27.2	30.8
	VSE-S*[61]	33.4	87.5	48.4	41.6	91.3	57.2	24.5	72.0	36.6	-	-	-
	AREN*†[50]	63.2	69.0	66.0	54.7	79.1	64.7	30.0	47.9	36.9	40.3	32.3	35.9
	DVBE	53.2	60.2	56.5	63.6	70.8	67.0	32.6	58.3	41.8	45.0	37.2	40.7
	DVBE*	64.4	73.2	**68.5**	62.7	77.5	**69.4**	37.9	55.9	**45.2**	44.1	41.6	**42.8**

ADAPTIVE AND GENERATIVE ZERO-SHOT LEARNING

Chou Y Y, Lin H T, Liu T L. Adaptive and generative zero-shot learning[C]//International conference on learning representations. 2021.

Abstract

CENTRAL SOUTH UNIVERSITY

1. seen/unseen 分类器：通过注意力机制提取视觉特征中与语义相关的信息并与语义信息结合以得到样本特定的语义信息，并通过 S2V 模块将其映射到视觉特征的维度以得到类级视觉特征，然后通过余弦相似度判断样本属于 seen 还是 unseen
2. seen 专家：余弦相似度 + softmax + 预测所有类别
3. unseen 专家：使用插值的方式生成虚拟类并对虚拟类进行学习

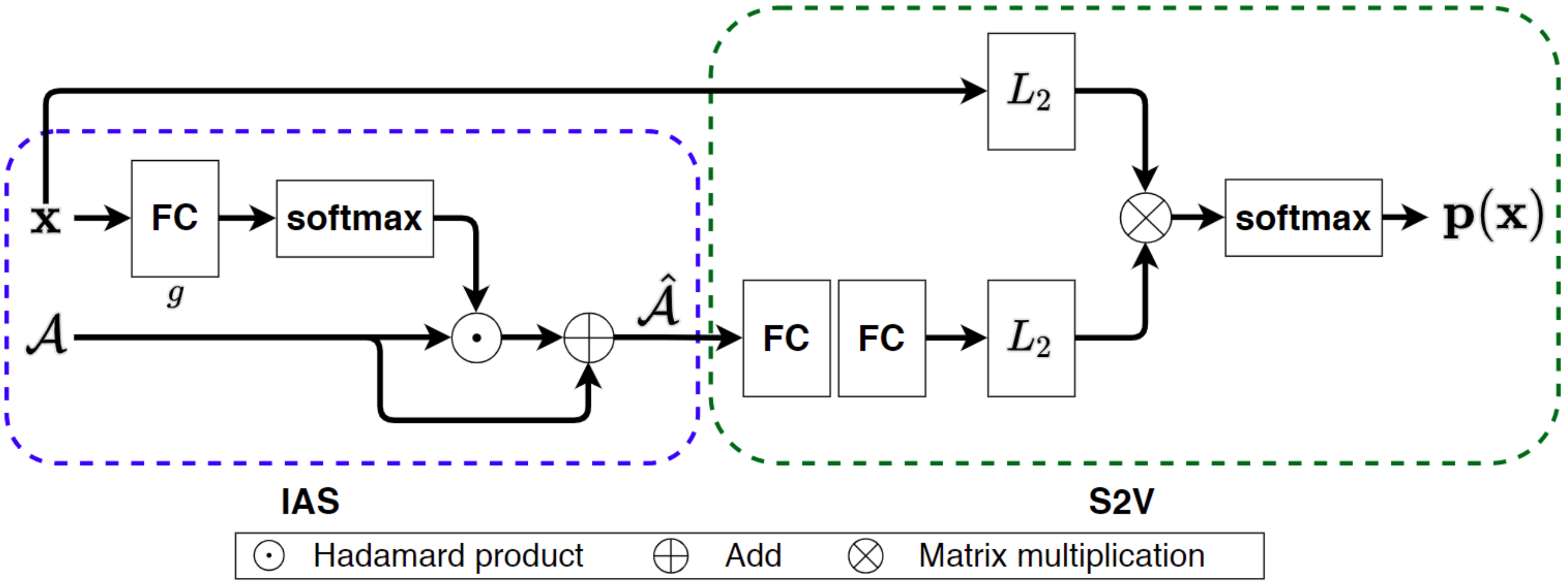


Figure 3: The architectures of IAS and S2V (FC: linear layer, L_2: L_2-normalization).

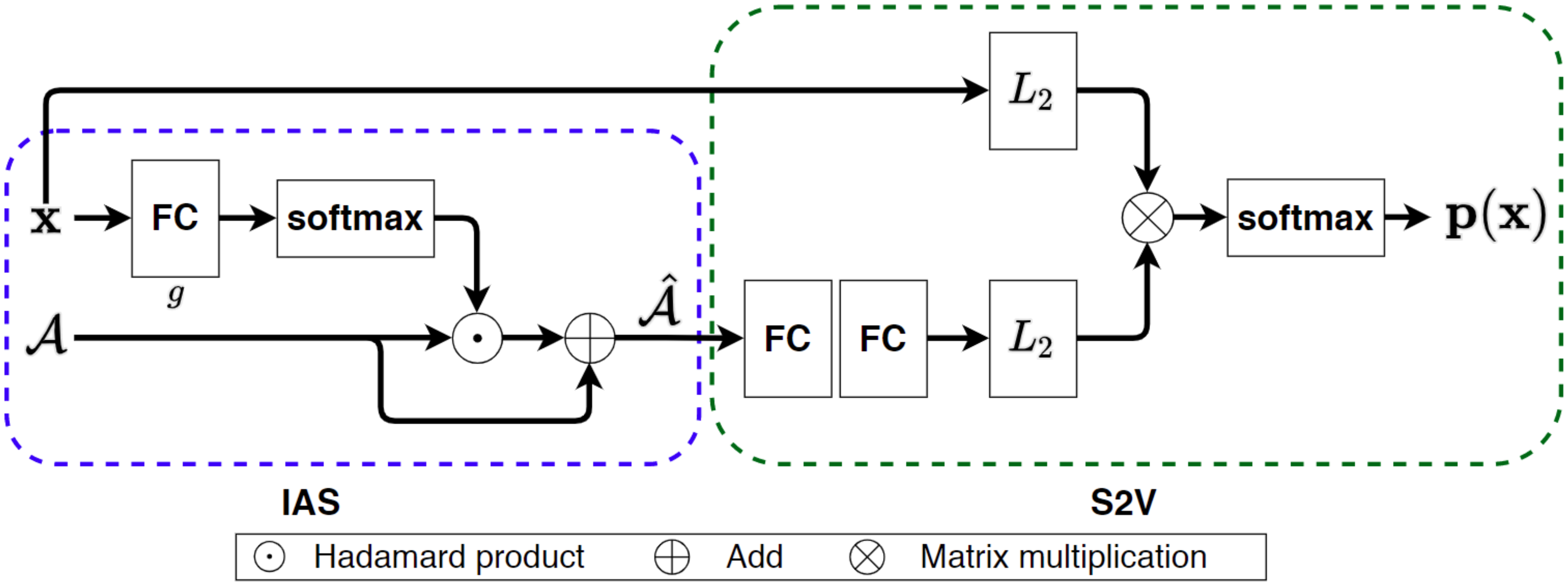
IAS + S2V

$$\hat{\mathcal{A}}(x) = \mathcal{A} + \mathcal{A} \odot \operatorname{softmax}(g(\mathbf{x}))$$

$$\mathcal{A} \in \mathbb{R}^{k \times n} \xmapsto[IAS]{x \in \mathbb{R}^d} \hat{\mathcal{A}}(x) \in \mathbb{R}^{k \times n} \longmapsto_{\mathrm{S2V}} \mathcal{V}(x) \in \mathbb{R}^{d \times n}$$

$$\mathbf{p}(\mathbf{x}) = f(\mathbf{x}, \mathcal{A}) = \operatorname{softmax}(\sigma \times \cos(\mathcal{V}(\mathbf{x}), \mathbf{x}))$$

$$\mathcal{L}_{CE} = -\sum_{\mathbf{x}} \log p_y(\mathbf{x})$$



分类

判别 seen/unseen

该论文根据余弦相似度的最大值是由 seen 还是 unseen 类别产生

- 若由 seen 产生，则直接分类为该类
- 若为 unseen 产生，则使用 unseen 专家进行分类

unseen 分类

对于 unseen 的分类学习，该论文使用元学习的思想进行训练，用混合插值的方式产生虚拟类

$$\mathbf{a}_v = \lambda \mathbf{a}_i + (1-\lambda)\mathbf{a}_j$$
$$\mathbf{x}_v = \lambda \mathbf{x}_i + (1-\lambda)\mathbf{x}_j$$

令 unseen 专家对虚拟类进行训练

Experiments

Central South University 中南大學 CENTRAL SOUTH UNIVERSITY

Table 1: GZSL results on four datasets. All methods in comparison utilize ResNet101 as the backbone for fairness. Notation "*" means the method fine-tunes the backbone to match the characteristic of datasets.

	AWA2				CUB				SUN				aPY			
	ZSL	GZSL			ZSL	GZSL			ZSL	GZSL			ZSL	GZSL		
	T1	U	S	H	T1	U	S	H	T1	U	S	H	T1	U	S	H
SP-AEN	58.5	23.3	90.9	37.1	55.4	34.7	70.6	46.6	59.2	24.9	38.6	30.3	24.1	13.7	63.4	22.6
DLFZRL	60.9	-	-	45.1	51.9	-	-	37.1	42.5	-	-	24.6	38.5	-	-	31.0
PSR	63.8	20.7	73.8	32.3	56.0	24.6	54.3	33.9	61.4	20.8	37.2	26.7	38.4	13.5	51.4	21.4
CDL	-	-	-	-	54.5	23.5	55.2	32.9	63.6	21.5	34.7	26.5	43.0	19.8	48.6	27.1
PQZSL	-	-	-	-	-	53.2	51.4	46.9	-	35.1	35.3	35.2	27.9	64.1	64.1	38.8
f-VAEGAN-D2*	70.3	57.1	76.1	65.2	72.9	63.2	75.6	68.9	65.6	50.1	37.8	43.1	-	-	-	-
LsrGAN	-	54.6	74.6	63.0	-	48.1	59.1	53.0	-	44.8	37.7	40.9	-	-	-	-
TF-VAEGAN*	73.4	55.5	83.6	66.7	74.3	63.8	79.3	70.7	**66.7**	41.8	51.9	46.3	-	-	-	-
OCD-CVAE	71.3	59.5	73.4	65.7	60.9	44.8	59.9	51.3	62.1	44.8	42.9	43.8	-	-	-	-
ZSML Softmax	76.1	58.9	74.6	65.8	69.6	60.0	52.1	55.7	60.2	-	-	-	**64.1**	36.3	46.6	40.9
GXE	71.1	56.4	81.4	66.7	54.4	47.4	47.6	47.5	62.6	36.3	42.8	39.3	38.0	26.5	74.0	39.0
E-PGN	73.4	52.6	83.5	64.6	72.4	52.0	61.1	56.2	-	-	-	-	-	-	-	-
Relation Net	64.2	30.0	93.4	45.3	55.6	38.1	61.1	47.0	-	-	-	-	-	-	-	-
Correlation Net	-	-	-	-	45.8	41.9	-	-	-	-	-	-	-	-	-	-
LFGAA+Hibrid	68.1	27.0	93.4	41.9	67.6	36.2	80.9	50.0	62.0	18.5	40.0	25.3	-	-	-	-
AREN*	66.9	54.7	79.1	64.7	72.5	63.2	69.0	66.0	60.6	40.3	32.3	35.9	39.2	30.0	47.9	36.9
COSMO	-	-	-	-	-	44.4	57.8	50.2	-	44.9	37.7	41.0	-	-	-	-
DVBE*	-	62.7	77.5	69.4	-	64.4	73.2	68.5	-	44.1	41.6	42.8	-	37.9	55.9	45.2
ours	73.8	65.1	78.9	71.3	57.2	41.4	49.7	45.2	63.3	29.9	40.2	34.3	41.0	35.1	65.5	**45.7**
ours*	**76.4**	69.0	86.5	**76.8**	**77.2**	69.2	76.4	**72.6**	66.2	50.5	43.1	**46.5**	43.7	36.2	58.6	44.8

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

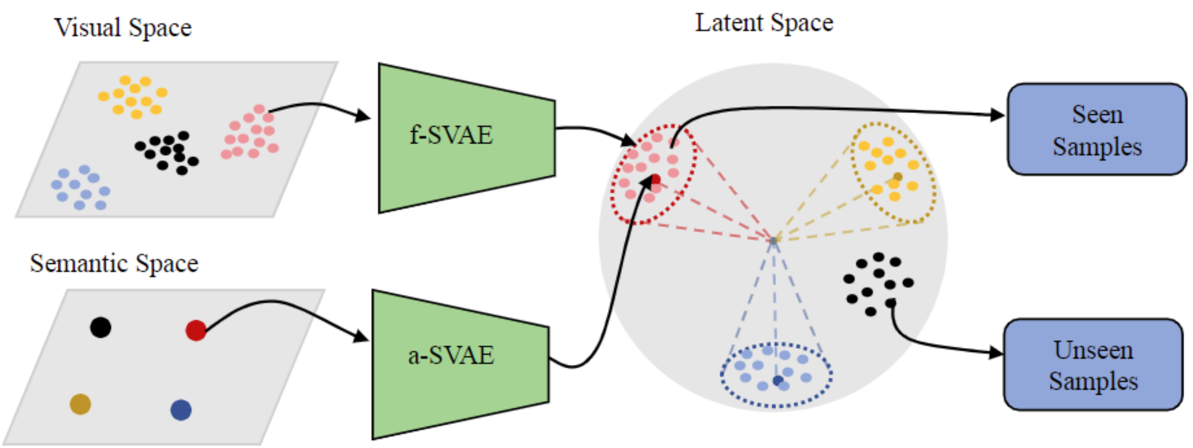


Fig. 1. The boundary based OOD classifier learns a bounded manifold for each seen class on a unit hyper-sphere (latent space). By using the manifold boundaries (dotted circles) and the centers (dark-colored dots), the unseen samples (black dots) can be separated from the seen samples (colored dots).

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：$p_\theta(\mathbf{z}|\mathbf{x})$

(a) Original (b) Autoencoder (c) $\mathcal{N}$-VAE (d) $\mathcal{N}$-VAE, $\beta = 0.1$ (e) $\mathcal{S}$-VAE

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

中南大學 CENTRAL SOUTH UNIVERSITY

- 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

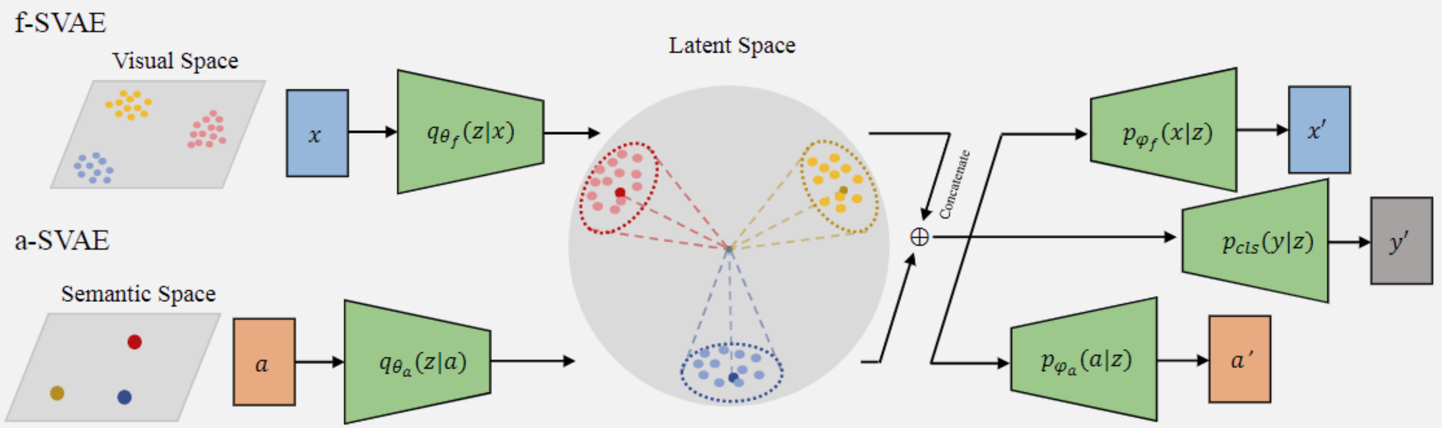


Fig. 2. Our model consists of two SVAEs, one for visual features and another for semantic attributes. By combining the objective functions of the two SVAEs with a cross-reconstruction loss and a classification loss, we train our model to align the latent distributions of visual features and semantic 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

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

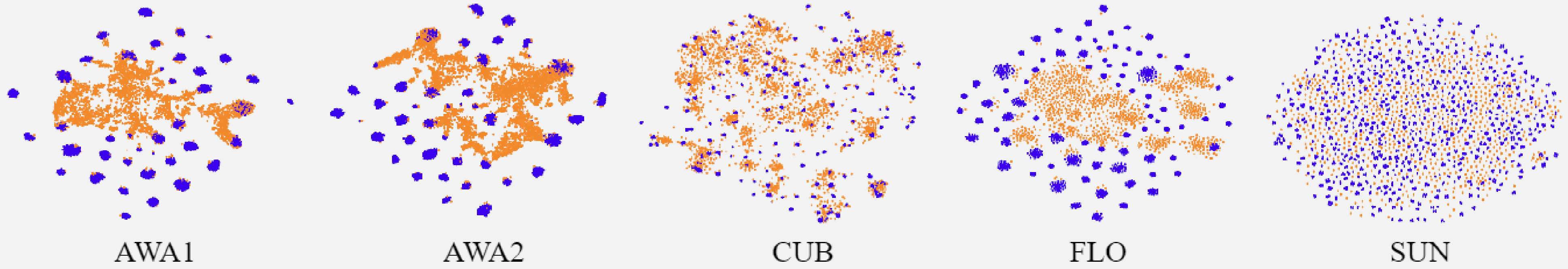


Fig. 4. The t-SNE visualization results for the learned latent space on the test sets of AWA1, AWA2, CUB, FLO and SUN. The blue dots represent the variables encoded from seen classes. The orange dots represent the variables encoded from unseen classes.

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

对比

	aPY			AWA1			AWA2			CUB		
	U	S	H	U	S	H	U	S	H	U	S	H
COSMO				52.8	80	63.6				44.4	57.8	50.2
DVBE	32.6	58.3	41.8				63.6	70.8	67.0	53.2	60.2	56.5
AGZSL	35.1	65.5	45.7				65.1	78.9	71.3	41.4	49.7	45.2
OOD				59.0	94.3	72.6	59.9	94.9	70.3	53.8	94.6	68.6

	FLO			SUN		
	U	S	H	U	S	H
COSMO				44.9	37.7	41.0
DVBE				45.0	37.2	40.7
AGZSL				29.9	40.2	34.3
OOD	61.9	91.7	73.9	57.8	95.1	71.9