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），Y^u（类别标签）

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

语义信息的不足

1. 事先定义的语义，可能缺少一些判别性的语义
    - 例如已有的语义有 `尾巴`、`羽毛` 等，但可能 `耳朵` 对于数据集十分有用
2. 部分语义对人来进行分类有用，但很难反映到视觉上

直接	间接	视觉无关，对人有帮助	意义不明
stripes 条纹	arctic 北极	hunter 食肉	newworld 新世界
furry 有毛	desert 沙漠	scavenger 食腐	oldworld 旧世界
longneck 长颈	forest 森林	agility 敏捷	
tail 尾巴	ocean 海洋	smart 聪明	

3. 部分类别间的语义差异小，导致很难区分
    - 例如 `猫` 和 `狗` 在语义上可能很接近，但在视觉上差异明显
4. 语义本身存在问题，不能很好地描述不同类别的属性

VGSE: Visually-Grounded Semantic Embeddings for Zero-Shot Learning

Reporter：陈思玉

2024.1.13

Xu W, Xian Y, Wang J, et al. Vgse: Visually-grounded semantic embeddings for zero-shot learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 9316-9325.

Author

CENTRAL SOUTH UNIVERSITY

WENJIA XU (许文嘉)

RESEARCH ASSOCIATE PROFESSOR,
PHD ADVISOR
BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS

Currently I am a research associate professor at School of Information and Communication Engineering, Beijing University of Posts and Telecommunications.

My research lies at the intersection of computer vision, natural language processing, and remote sensing, which covers a wide range of topics including remote sensing scene classification, few-shot learning, image captioning, explainable artificial inteligience.

I got my PhD degree from University of Chinese Academy of Sciences, advised by Prof. Yirong Wu. During my PhD, I'm also co-supervised by Prof. Zeynep Akata and Prof. Bernt Schiele at Max Plank Institute for Informatics. I got my Bachelar's degree with horner from Beijing Institute of Technology at 2016. During Jan.2016 - Jul.2016, I finished my bachelor thesis under the supervision of Prof. Tobias Oechtering at KTH.

- 04/2022: Our paper about Zero-Shot Learning is accepted by IJCAI 2022.
- 03/2022: Our paper about Zero-Shot Learning is accepted by IJCV.
- 03/2022: Our paper about distinctive image captioning is accepted by TPAMI.
- 03/2022: Our paper about Zero-Shot Learning is accepted by CVPR 2022.
- 01/2022: I am awarded the National Scholarship for Graduate Students by MOE (Ministry Of Education).
- 10/2021: Our paper about the intersaction of XAI and gaze is accepted by BMVC 2021.
- 06/2021: One paper about distinctive image captioning is accepted by ACM MM 2021 as Oral.
- 09/2020: One paper about Zero-Shot Learning is accepted by NeurIPS 2020.
- 07/2020: One paper about distinctive image captioning is accepted by ECCV 2020 as Oral.
- 07/2020: Two workshop papers accepted by ECCV 2020.

VGSE

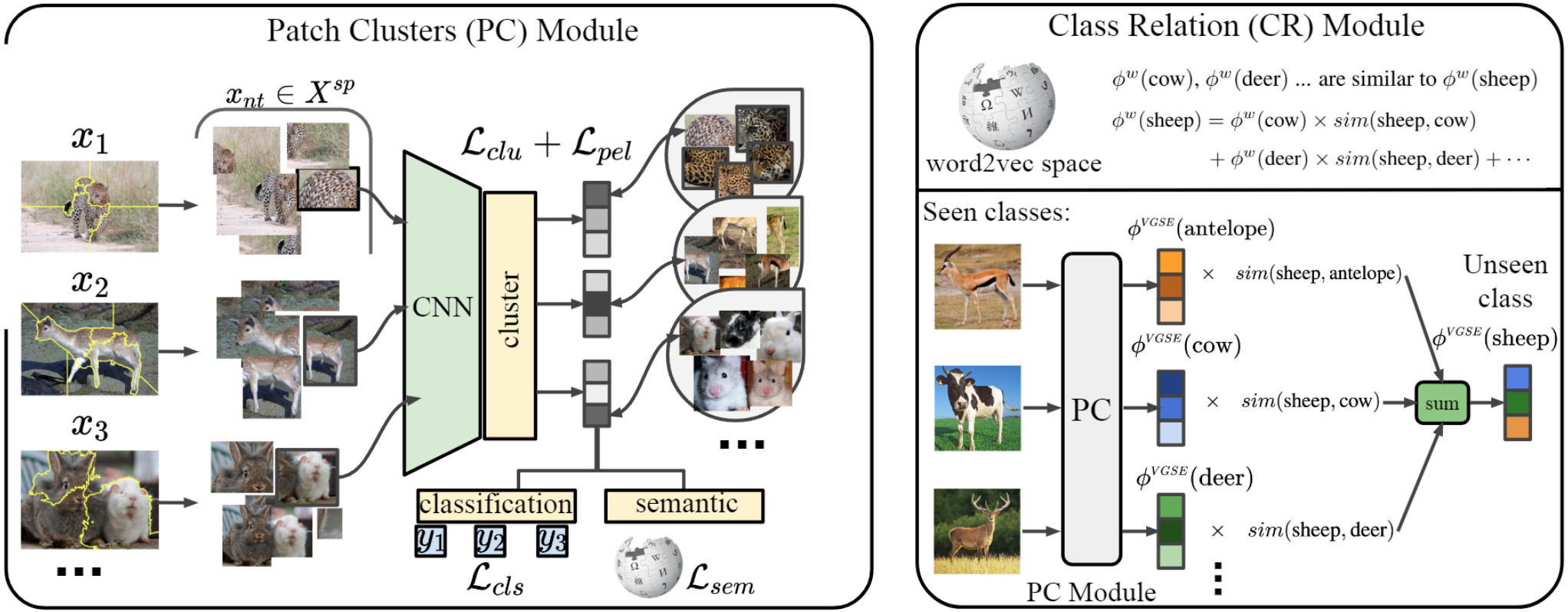


Figure 2. Our visually-grounded semantic embedding network consists of two modules. The Patch Clustering (PC) module learns clusters from patch images, and predicts semantic embeddings for seen classes with their images. The Class Relation (CR) module predicts the unseen class embeddings $\phi^{VGSE}(y_m)$ using unseen and seen class relations learned from external knowledge, e.g., word2vec. For instance, the embedding for unseen class *sheep* is predicted using the semantic embeddings of the seen classes, e.g., *antelope*, *cow*, *deer*, and so on.

VGSE

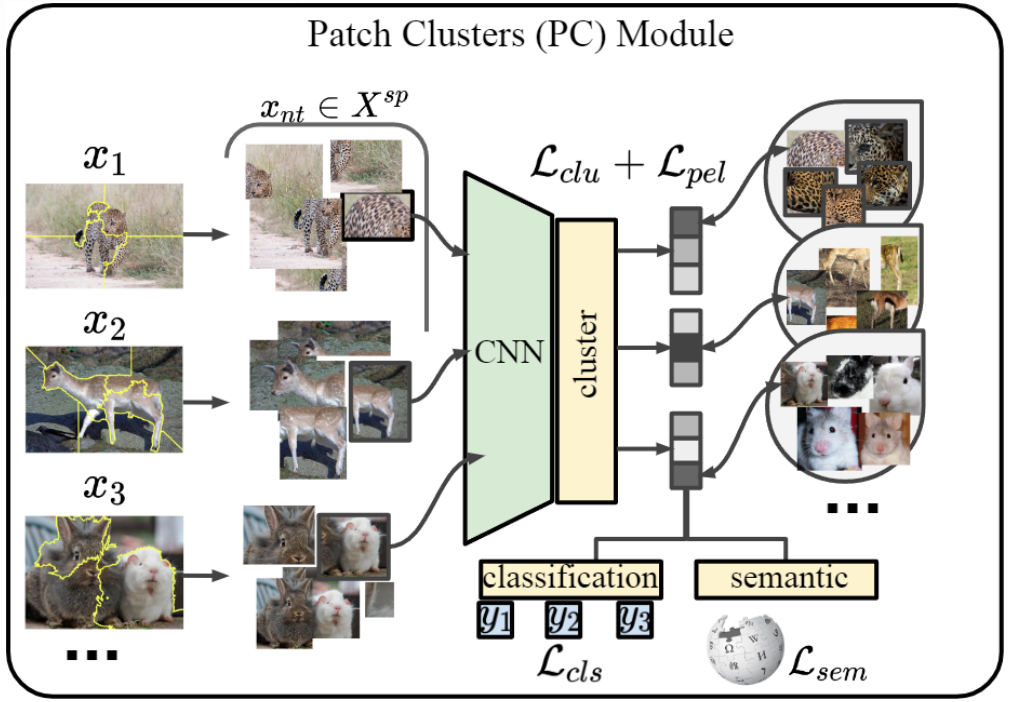
Patch Clusters：训练

1. 图像切块：紧凑分水岭分割算法 $N_t = 9$
   - $|X^{sp}| = N_s N_t$
2. 图像块聚类：聚类为 K 类 $a_t = H \circ \theta(x_t)$
   - 聚类损失：

     $\mathcal{L}_{clu} = -\sum_{x_t \in X^{sp}} \sum_{x_i \in X_b^{sp}} \log(a_t^T a_i)$
     - $X_b^{sp} \rightarrow \|\theta(x_t) - \theta(x_i)\|_2$
   - 防止所有块被分配到相同簇：

     $\mathcal{L}_{pel} = \sum_{k=1}^{K} \bar{a}_t^k \log \bar{a}_t^k$

     $\bar{a}_t^k = \frac{1}{N_s N_t} \sum_{x_t \in X^{sp}} a_t^k$

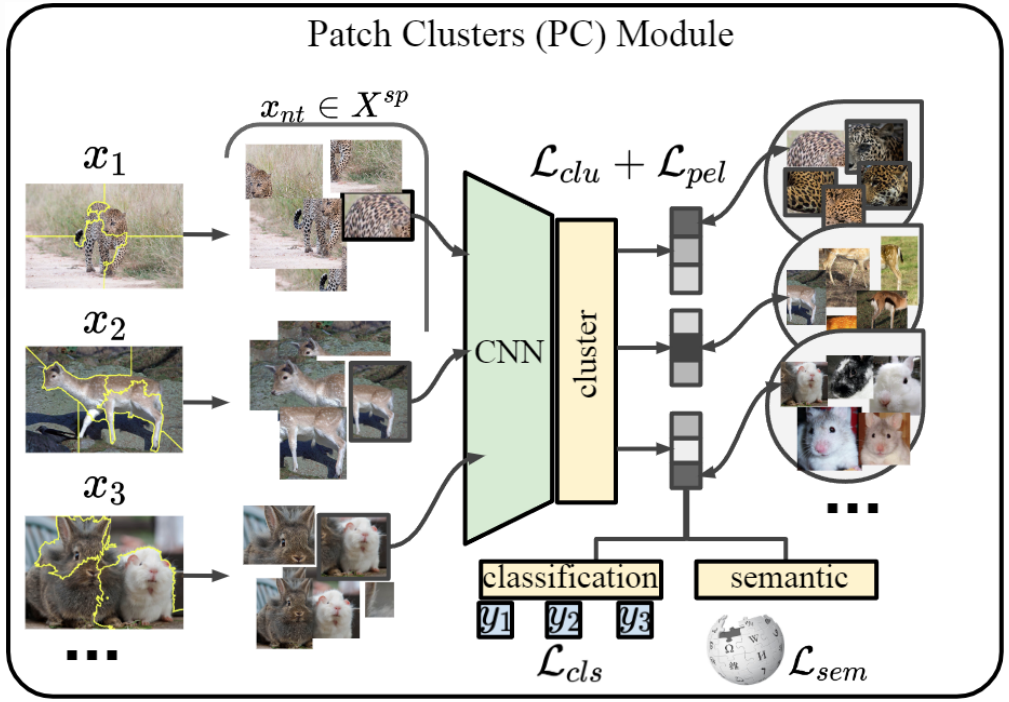


Van Gansbeke W, Vandenhende S, Georgoulis S, et al. Scan: Learning to classify images without labels[C]//European conference on computer vision. Cham: Springer International Publishing, 2020: 268-285.

VGSE

Patch Clusters：训练

3. 分类损失：
   - $\mathcal{L}_{cls} = -\log \frac{\exp(p(y_n|x_t))}{\sum_{\hat{y}\in Y^s}\exp(p(\hat{y}|x_t))}$
4. 语义相关性：
   - $\mathcal{L}_{sem} = \|S \circ a_t - a^{ori}\|_2$
   - *?：如何保证新的语义不与已有的语义重复*



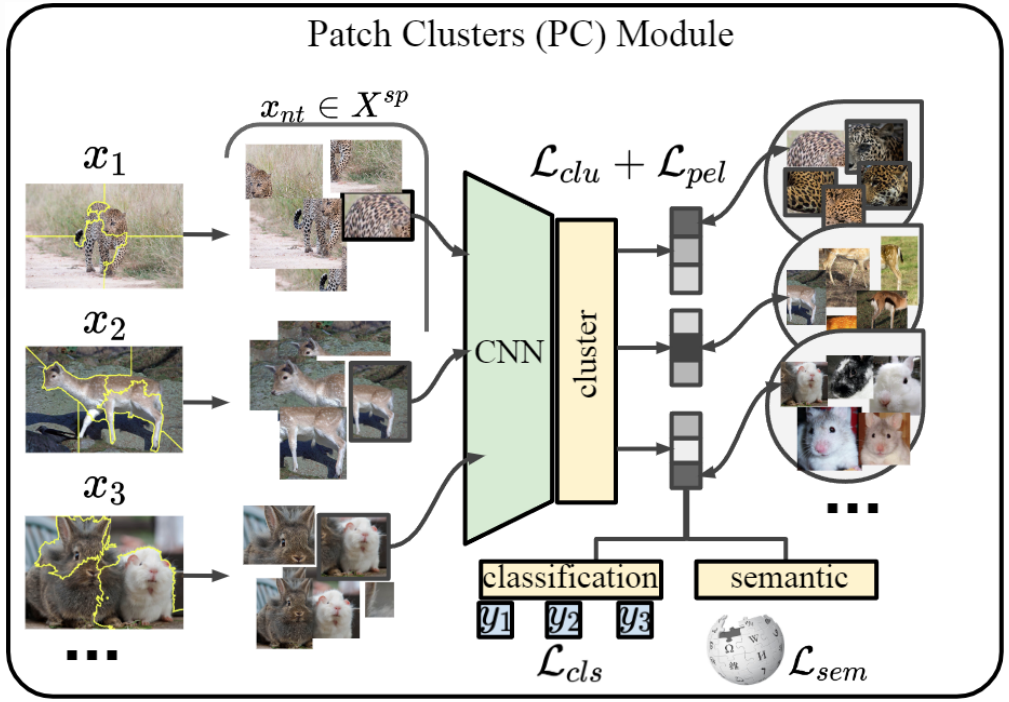
总损失：

$$\mathcal{L} = \mathcal{L}_{clu} + \lambda\mathcal{L}_{pel} + \beta\mathcal{L}_{cls} + \gamma\mathcal{L}_{sem}$$

Patch Clusters：构建可见类的增强语义

1. 可见类样本的增强语义：
   - $a^{ins} = \frac{1}{N_t} \sum_{t=1}^{N_t} a_t$

2. 可见类的增强语义：
   - $a^{sc} = \frac{1}{N_c^s} \sum_{j \in Y_c^s} a_j^{ins}$



VGSE

Class Relation：构建未见类的增强语义

1. WAvg：相似可见类 Y_b^s 的组合

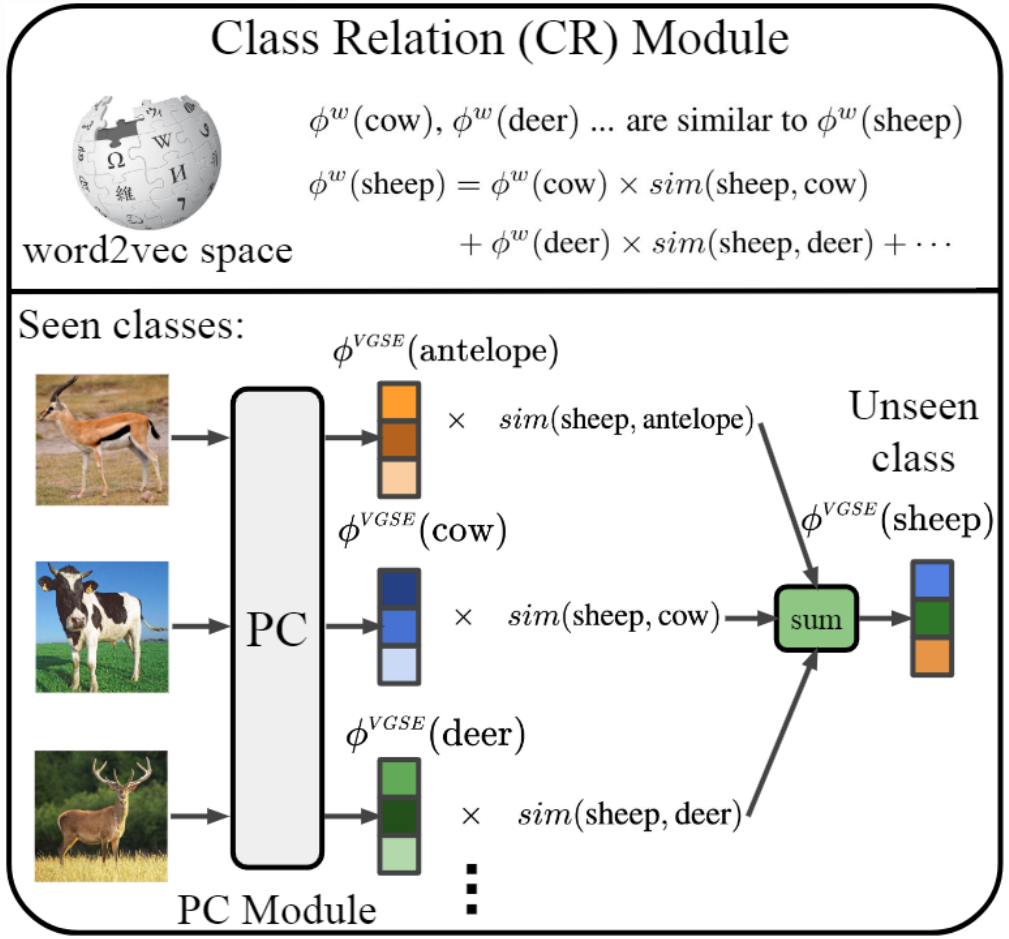
$$a^{uc} = \frac{1}{|Y_b^s|} \sum_{\tilde{a}^{ori} \in Y_b^s} sim(a^{ori}, \tilde{a}^{ori}) \cdot \tilde{a}^{sc}$$

$$sim(a^{ori}, \tilde{a}^{ori}) = \exp(-\eta \| a^{ori} - \tilde{a}^{ori} \|_2)$$

2. SMO：Similarity Matrix Optimization

$$\min_r \| a^{ori} - r^T A^{ori,s} \|_2$$

$$\text{s.t.} \quad \alpha < r < 1 \quad and \quad \sum_{i=1}^{|Y^s|} r_i = 1$$



Experiments

Comparison with SOTA

	ZSL Model	Semantic Embeddings	Zero-Shot Learning			Generalized Zero-Shot Learning								
			AWA2	CUB	SUN	AWA2			CUB			SUN		
			T1	T1	T1	u	s	H	u	s	H	u	s	H
Generative	CADA-VAE [43]	w2v [31]	49.0	22.5	37.8	38.6	60.1	47.0	16.3	39.7	23.1	26.0	28.2	27.0
		VGSE-SMO (Ours)	52.7	24.8	40.3	46.9	61.6	53.9	18.3	44.5	25.9	**29.4**	29.6	29.5
	f-VAEGAN-D2 [61]	w2v [31]	58.4	32.7	39.6	46.7	59.0	52.2	23.0	44.5	30.3	25.9	33.3	29.1
		VGSE-SMO (Ours)	61.3	**35.0**	**41.1**	45.7	66.7	54.2	**24.1**	45.7	**31.5**	25.5	**35.7**	**29.8**
Non-Generative	SJE [2]	w2v [31]	53.7	14.4	26.3	39.7	65.3	48.8	13.2	28.6	18.0	19.8	18.6	19.2
		VGSE-SMO (Ours)	62.4	26.1	35.8	46.8	72.3	56.8	16.4	44.7	28.3	28.7	25.2	26.8
	GEM-ZSL [28]	w2v [31]	50.2	25.7	-	40.1	80.0	53.4	11.2	**48.8**	18.2	-	-	-
		VGSE-SMO (Ours)	58.0	29.1	-	49.1	78.2	60.3	13.1	43.0	20.0	-	-	-
	APN [62]	w2v [31]	59.6	22.7	23.6	41.8	75.0	53.7	17.6	29.4	22.1	16.3	15.3	15.8
		VGSE-SMO (Ours)	**64.0**	28.9	38.1	**51.2**	**81.8**	**63.0**	21.9	45.5	29.5	24.1	31.8	27.4

Table 1. Comparing our VGSE-SMO, with w2v semantic embedding over state-of-the-art ZSL models. In ZSL, we measure Top-1 accuracy (**T1**) on unseen classes, in GZSL on seen/unseen (**s/u**) classes and their harmonic mean (**H**). Feature Generating Methods, i.e., f-VAEGAN-D2, and CADA-VAE generating synthetic training samples, and SJE, APN, GEM-ZSL using only real image features.

Experiments

Comparison with SOTA

Semantic Embeddings	External knowledge	Zero-shot learning AWA2	CUB	SUN
w2v [31]	w2v	58.4	32.7	39.6
ZSLNS [39]	T	57.4	27.8	-
GAZSL [67]	T	-	34.4	-
Auto-dis [3]	T	52.0	-	-
CAAP [5]	T and H	55.3	31.9	35.5
VGSE-SMO (Ours)	w2v	**61.3** $\pm$ 0.3	**35.0** $\pm$ 0.2	**41.1** $\pm$ 0.3

Table 2. Comparing with state-of-the-art methods for learning semantic embeddings with less human annotation (T: online textual articles, H: human annotation) using same image features and ZSL model (f-VAEGAN-d2 [61]).

Experiments

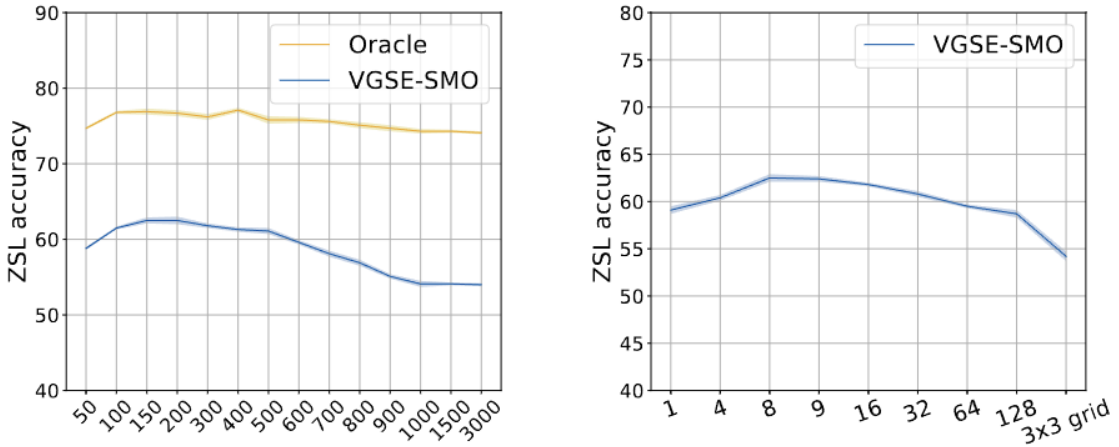
Ablation Study

Semantic Embeddings	Zero-shot learning		
	AWA2	CUB	SUN
k-means-SMO	54.5 ± 0.4	15.0 ± 0.5	25.2 ± 0.4
ResNet-SMO	55.3 ± 0.2	15.4 ± 0.1	25.1 ± 0.1
$\mathcal{L}_{clu} + \mathcal{L}_{pel}$ (baseline + SMO)	56.6 ± 0.2	16.7 ± 0.2	26.3 ± 0.3
$+\mathcal{L}_{cls}$	61.2 ± 0.1	23.7 ± 0.2	30.5 ± 0.2
$+\mathcal{L}_{sem}$ (VGSE-SMO)	$\mathbf{62.4} \pm 0.3$	$\mathbf{26.1} \pm 0.3$	$\mathbf{35.8} \pm 0.2$
VGSE-WAvg	57.7 ± 0.2	25.8 ± 0.3	35.3 ± 0.2

Table 3. Ablation study over the PC module reporting ZSL T1 on AWA2, CUB, and SUN (mean accuracy and std over 5 runs). The baseline is the PC module with the cluster loss $\mathcal{L}_{clu}$ and $\mathcal{L}_{pel}$. Our full model VGSE-SMO is trained with two additional losses $\mathcal{L}_{cls}$, $\mathcal{L}_{sem}$. Two kinds of semantic embeddings learned from k-means clustering and pretrained ResNet are listed below for comparison.

Experiments

Cluster Number and Patch Number



(a) cluster number D_v (b) patch number N_t

Figure 3. (a) Influence of the cluster number $D_v = 50, \ldots, 3000$. In the oracle setting, we feed unseen classes images to the PC module to predict unseen semantic embeddings. (b) Influence of the patch number N_t we used per image with the watershed segmentation for obtaining our VGSE−SMO class embeddings. $N_t = 1$ uses the whole image (no patches). "3×3 grid" crops the image into 9 square patches. Both plots report ZSL accuracy with SJE model trained on AWA2 dataset (mean and std over 5 runs).

Experiments

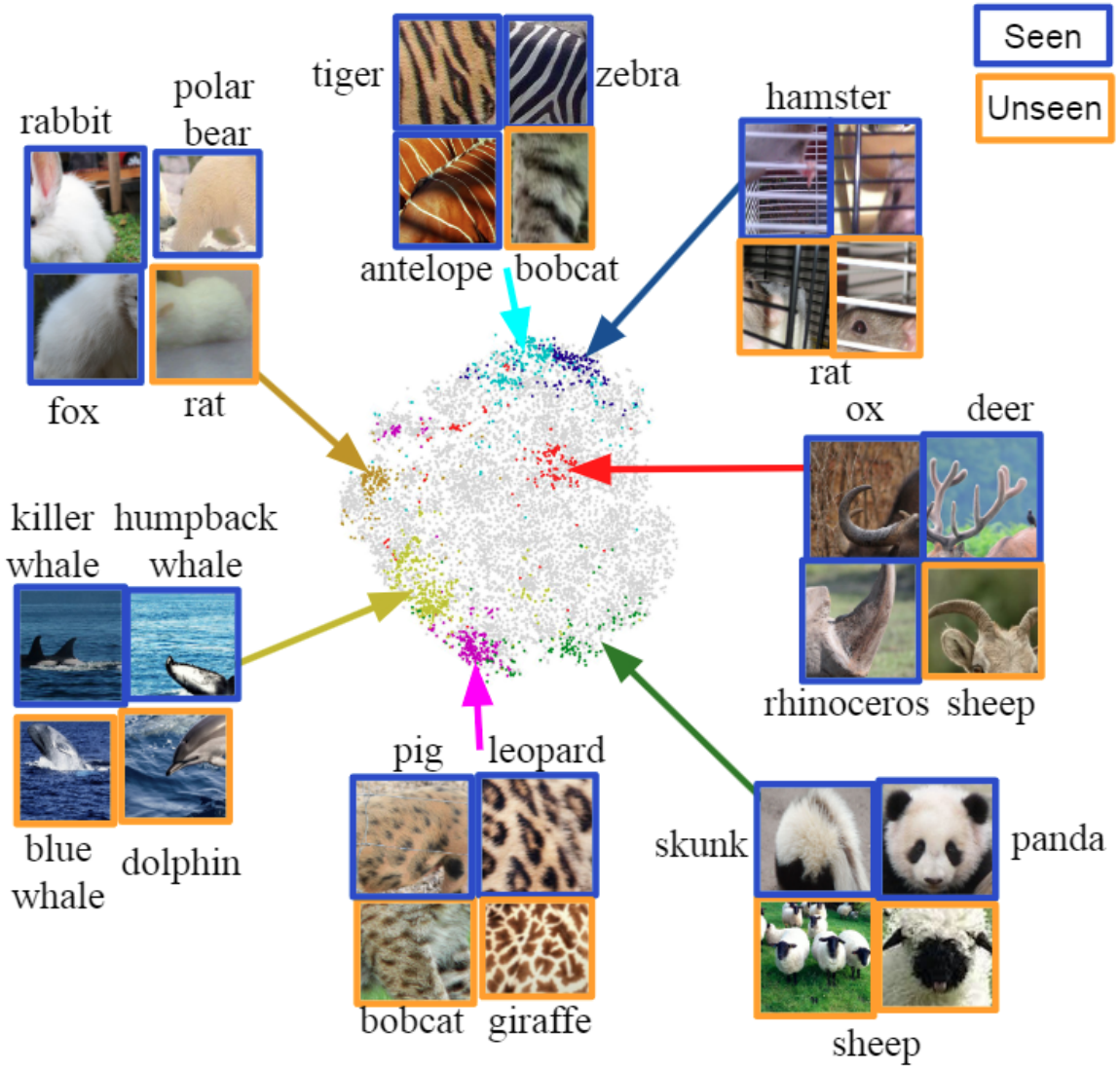
Semantic Type

Semantic Embeddings	AWA2		CUB	
	T1	H	T1	H
w2v [31]	53.7 ± 0.2	48.8 ± 0.1	14.4 ± 0.3	18.0 ± 0.2
VGSE-SMO (w2v)	62.4 ± 0.1	56.8 ± 0.1	26.1 ± 0.2	28.3 ± 0.1
glove [38]	38.8 ± 0.2	38.7 ± 0.3	19.3 ± 0.2	13.4 ± 0.1
VGSE-SMO (glove)	46.5 ± 0.1	46.0 ± 0.1	25.2 ± 0.3	27.1 ± 0.2
fasttext [7]	47.7 ± 0.1	44.6 ± 0.3	-	-
VGSE-SMO (fasttext)	51.9 ± 0.2	53.2 ± 0.1	-	-
Attribute	62.8 ± 0.1	62.6 ± 0.3	56.4 ± 0.2	49.4 ± 0.1
VGSE-SMO (Attribute)	66.7 ± 0.1	64.9 ± 0.1	56.8 ± 0.1	50.9 ± 0.2

Table 4. Evaluating the external knowledge, i.e., word embeddings w2v [31], glove [38], fasttext [7], and the human annotated attributes, for our VGSE-SMO embeddings, e.g., VGSE-SMO (glove) indicates that CR module is trained with glove embedding. **T1**: top-1 accuracy in ZSL, **H**: harmonic mean in GZSL trained with SJE [2] on AWA2, and CUB (std over 5 runs).

Experiments

Qualitative Results



Evolving Semantic Prototype Improves Generative Zero-Shot Learning

Reporter：陈思玉

2024.1.13

Chen S, Hou W, Hong Z, et al. Evolving Semantic Prototype Improves Generative Zero-Shot Learning[J]. arXiv preprint arXiv:2306.06931, 2023.

Author

Shiming Chen 陈使明

Postdoctoral Research Fellow

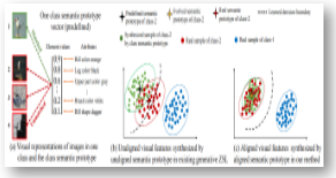
Email: gchenshiming **at** gmail **dot** com

Short Bios

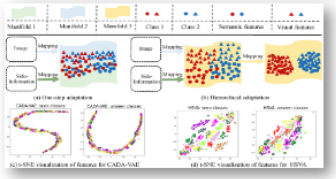
I received my Ph.D. degree at Huazhong University of Science and Technology in Dec. 2022, advised by Prof. Xinge You and worked closely Prof. Ling Shao. My current research interests span computer vision and machine learning with a series of topics, such as ***zero-shot learning***, ***generative modeling and learning***, and ***visual-and-language learning***.

Author

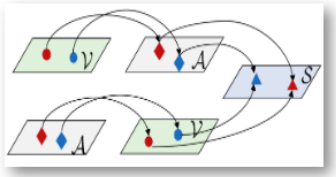
Conference Papers



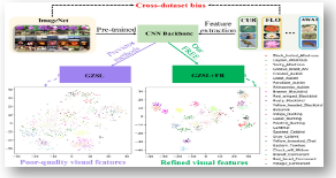
Evolving Semantic Prototype Improves Generative Zero-Shot Learning. ***[PDF]*** *[arXiv]*
Shiming Chen*, Wenjin Hou, Ziming Hong, Xiaohan Ding, Yibing Song, Xinge You, Tongliang Liu, Kun Zhang.*
The Fortieth International Conference on Machine Learning (***ICML****), 2023.* ***(CCF Rank-A)***



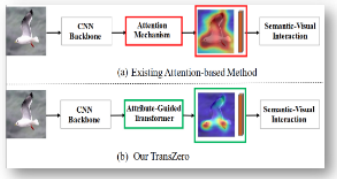
HSVA: Hierarchical Semantic-Visual Adaptation for Zero-Shot Learning. ***[PDF] [arXiv] [Code]***
Shiming Chen*, Guo-Sen Xie, Qinmu Peng, Yang Liu, Baigui Sun, Hao Li, Xinge You, Ling Shao.*
Annual Conference on Neural Information Processing Systems (***NeurIPS****), 2021: 16622-16634.*
(CCF Rank-A)



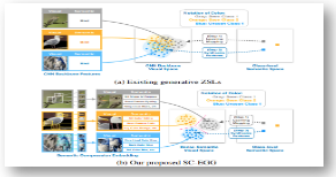
MSDN: Mutually Semantic Distillation Network for Zero-Shot Learning. ***[PDF] [arXiv] [Code]***
Shiming Chen*, Ziming Hong, Guo-Sen Xie, Wenhan Yang, Qinmu Peng, Kai Wang, Jian Zhao, Xinge You.*
IEEE Conference on Computer Vision and Pattern Recognition (***CVPR****), 2022: 7612-7621.* ***(CCF Rank-A)***



FREE: Feature Refinement for Generalized Zero-shot Learning. ***[PDF] [arXiv] [Code]***
Shiming Chen*, Wenjie Wang, Beihao Xia, Qinmu Peng, Xinge You, Feng Zheng, Ling Shao.*
IEEE International Conference on Computer Vision (***ICCV****), 2021: 1106-1112.* ***(CCF Rank-A)***



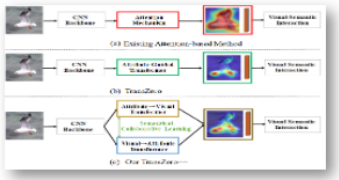
TransZero: Attribute-guided Transformer for Zero-Shot Learning. ***[PDF] [arXiv] [Code]***
Shiming Chen*[*], Ziming Hong[*], Yang Liu, Guo-Sen Xie, Baigui Sun, Hao Li, Qinmu Peng, Ke Lu, Xinge You.*
Thirty-Sixth AAAI Conference on Artificial Intelligence (***AAAI****), 2022: 330-338.* ***(CCF Rank-A)***



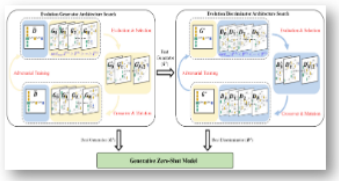
Semantic Compression Embedding for Generative Zero-Shot Learning.
Ziming Hong,* ***Shiming Chen****[*#], Guo-Sen Xie, Wenhan Yang, Jian Zhao, Yuanjie Shao, Qinmu Peng, Xinge You*
The 31th International Joint Conference on Artificial Intelligence (***IJCAI****), 2022: 956-963.* ***(CCF Rank-A)***

Author

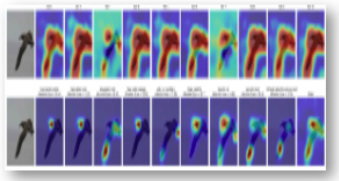
Journal Papers



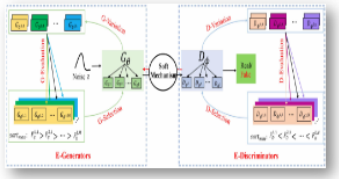
TransZero++: Cross Attribute-guided Transformer for Zero-Shot Learning. ***[Project Page]*** ***[arXiv]*** ***[PDF]***
Shiming Chen*, Ziming Hong, Wenjin Hou, Guo-Sen Xie, Yibing Song, Jian Zhao, Xinge You, Shuicheng Yan, Ling Shao.*
IEEE Transactions on Pattern Analysis and Machine Intelligence (***TPAMI****), 45(11):12844-12861, 2023.* ***(SCI, IF=24.314, CCF Rank-A)***



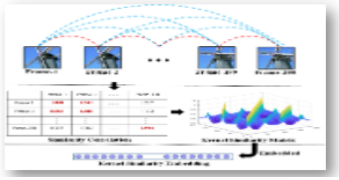
EGANS: Evolutionary Generative Adversarial Network Search for Zero-Shot Learning. ***[PDF]*** ***[arXiv]***
Shiming Chen*, Shuhuang Chen, Wenjin Hou,Weiping Ding, Xinge You.*
IEEE Transactions on Evolutionary Computation (***TEC****), in press, 2023.* ***(SCI, IF=14.3, CCF Rank-B)***



GNDAN: Graph Navigated Dual Attention Network for Zero-Shot Learning. ***[Code]*** ***[PDF]***
Shiming Chen*, Ziming Hong, Guo-Sen Xie, Xinge You, Weiping Ding and Ling Shao.*
IEEE Transactions on Neural Networks and Learning Systems (***TNNLS****), in press, 2022.* ***(SCI, IF=14.255, CCF Rank-B)***



CDE-GAN: Cooperative Dual Evolution Based Generative Adversarial Network. ***[PDF]*** ***[arXiv]***
Shiming Chen*, Wenjie Wang, Beihao Xia, Xinge You, Qinmu Peng, Zehong Cao, Weiping Ding.*
IEEE Transactions on Evolutionary Computation (***TEC****), 25:986-1000, 2021.* ***(SCI, IF=14.3, CCF Rank-B)***



Kernelized Similarity Learning and Embedding for Dynamic Texture Synthesis. ***[Code]*** ***[arXiv]***
Shiming Chen*, Peng Zhang, Guo-sen Xie, Zehong Cao, Qinmu Peng, Wei Yuan, Xinge You.*
IEEE Transactions on Systems, Man and Cybernetics: Systems (***TSMCA****), 53(2):824-837, 2023.*
(SCI, IF=11.471)

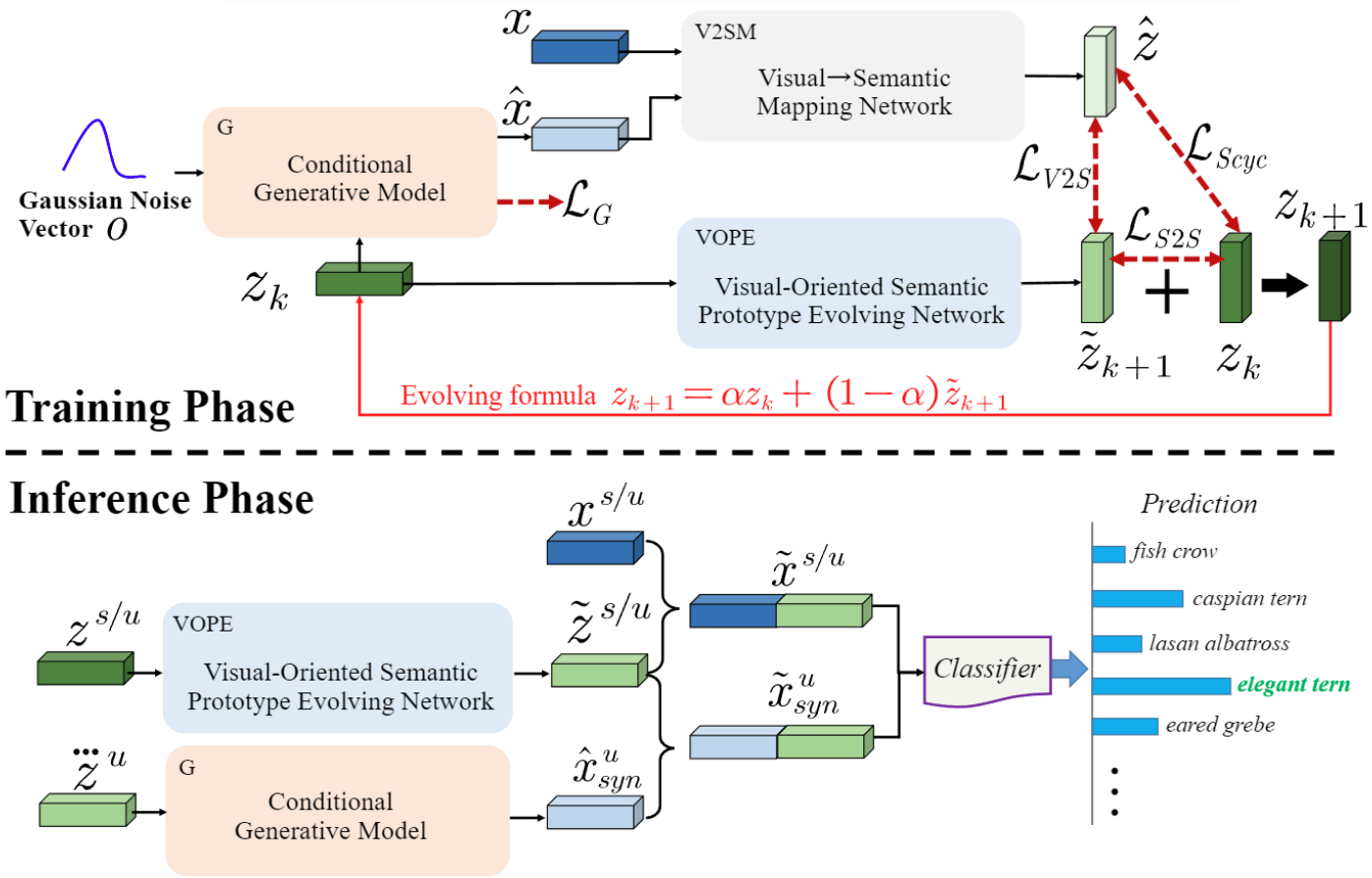


Figure 2. Dynamic semantic prototype evolvement (DSP). During training, we use V2SM and VOPE networks. V2SM maps sample features into semantic prototypes. VOPE maps dynamic semantic prototypes from the k-th step to the $(k+1)$-th step for evolvement. Based on the prototype z_k in the k-th step, we use the conditional generator to synthesize sample features $\hat{x}$ and map $\hat{x}$ to prototype $\hat{z}$ via V2SM. We use $\hat{z}$ to supervise VOPE output $\hat{z}_{k+1}$ during evolvement, which brings semantics from sample features. The V2SM and VOPE are jointly trained with the generative model. During inference, we use VOPE to map one input prototype $z^{s/u}$ to its evolved form $\tilde{z}^{s/u}$ for both seen and unseen classes. Besides, we take $\ddot{z}^u = \alpha \cdot z^u + (1-\alpha) \cdot \tilde{z}^u$ as the generator input to synthesize sample features $\hat{x}^u_{syn}$ for unseen classes. Then, we concatenate sample features with their $\tilde{z}^{s/u}$ for semantic enhancement during ZSL classification.

V2SM

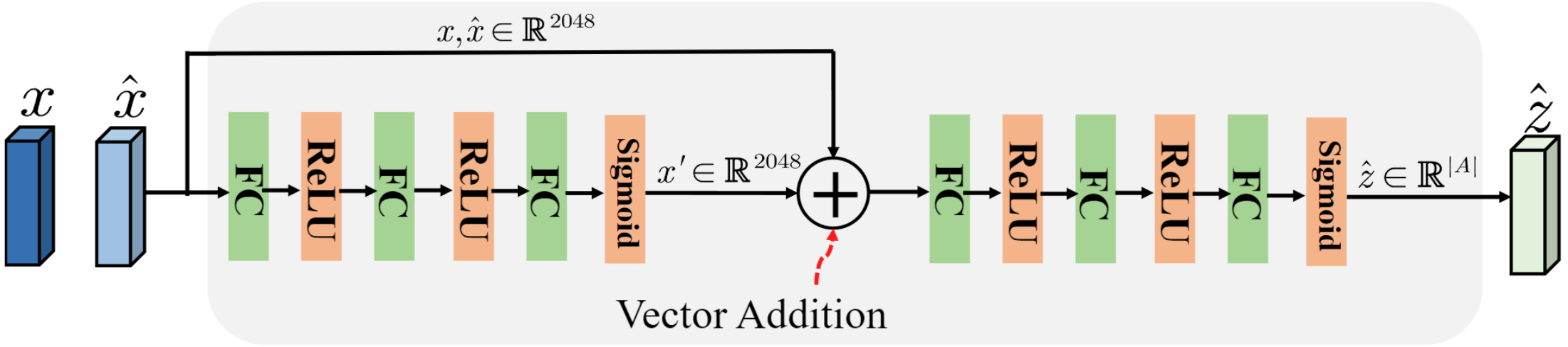


Figure 5. Network details of V2SM.

$$\hat{z}_{real} = V2SM(x) \quad or \quad \hat{z}_{syn} = V2SM(\hat{x})$$

$$\mathcal{L}_{Scyc} = \mathbb{E}[\|\hat{z}_{real} - z_k\|_1] + \mathbb{E}[\|\hat{z}_{syn} - z_k\|_1]$$

DSP

VOPE

$$\tilde{z}_{k+1} = VOPE(z_k) \quad z_0 = z$$

$$\mathcal{L}_{V2S} = \mathbb{E}[1 - \cos(\hat{z}, \tilde{z}_{k+1})]$$

$$\mathcal{L}_{S2S} = \mathbb{E}[\|\tilde{z}_{k+1} - z_k\|_1]$$

$$z_{k+1} = \alpha z_k + (1 - \alpha)\tilde{z}_{k+1}$$

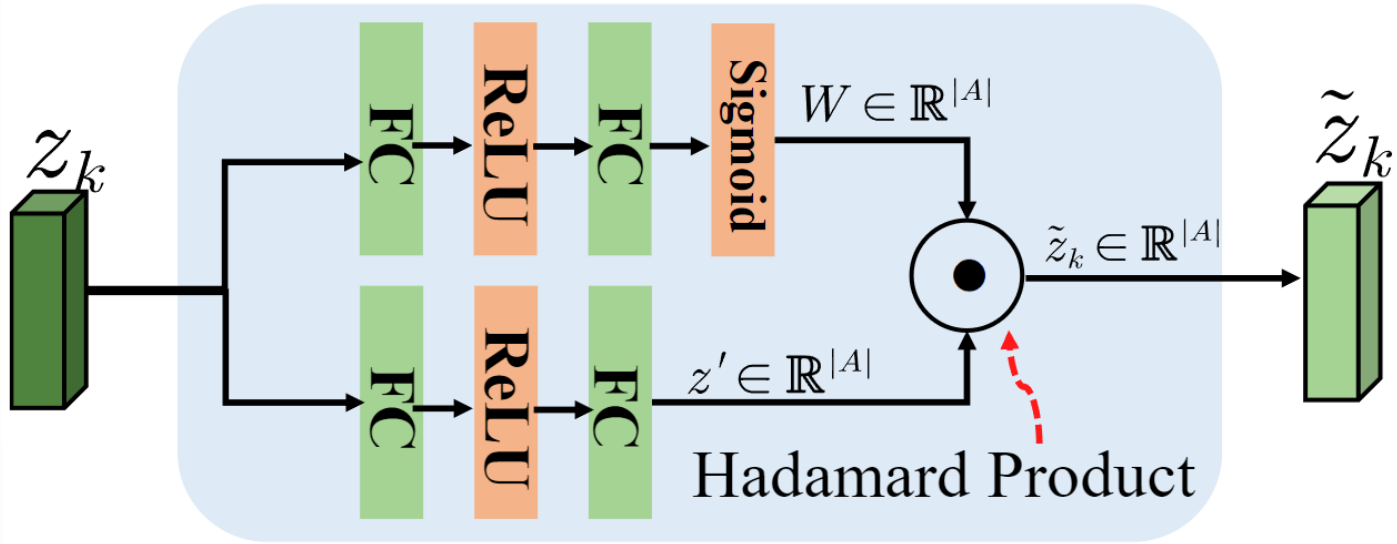


Figure 6. Network details of VOPE.

总损失：

$$\mathcal{L}_{total} = \mathcal{L}_G + \lambda_{Scyc}\mathcal{L}_{Scyc} + \lambda_{V2S}\mathcal{L}_{V2S} + \lambda_{S2S}\mathcal{L}_{S2S}$$

Inference Phase

1. 生成未见类的动态语义原型

$$z_0^{c^u} = z^{c^u}$$

$$z_{k+1}^{c^u} = \alpha z_k^{c^u} + (1-\alpha)\tilde{z}_{k+1}^{c^u}$$

2. 生成未见类的视觉样本

$$\hat{x}_{syn}^u = G(o, z_{k+1}^{c^u})$$

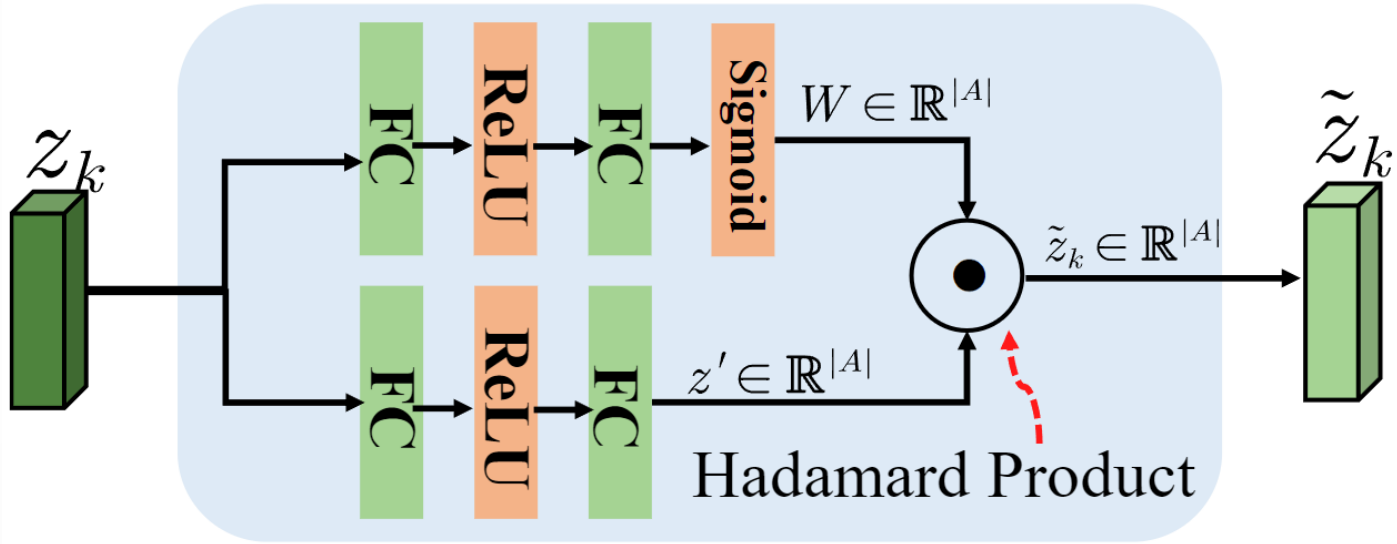


Figure 6. Network details of VOPE.

3. 拼接视觉特征和动态语义原型，以训练分类器

Seen Classes : $\tilde{x}_{tr}^s = x_{tr}^s \oplus \tilde{z}_{k+1}^s$; $\tilde{x}_{te}^s = x_{te}^s \oplus \tilde{z}_{k+1}^s$

Unseen Classes : $\tilde{x}_{syn}^u = \tilde{x}_{syn}^u \oplus \tilde{z}_{k+1}^u$; $\tilde{x}_{te}^u = x_{te}^u \oplus \tilde{z}_{k+1}^u$

Experiments

Comparison with SOTA

Table 1. State-of-the-art comparisons on CUB, SUN, AWA2, and FLO under GZSL settings. Embedding-based methods are categorized as ♣, and generative methods are categorized as ♠. The best and second-best results are marked in **Red** and **Blue**, respectively.

	Methods	Venue	CUB			SUN			AWA2			FLO		
			U	S	H	U	S	H	U	S	H	U	S	H
♣	SGMA (Zhu et al., 2019)	NeurIPS'19	36.7	71.3	48.5	–	–	–	37.6	87.1	52.5	–	–	–
	AREN (Xie et al., 2019)	CVPR'19	38.9	**78.7**	52.1	19.0	38.8	25.5	15.6	**92.9**	26.7	–	–	–
	CRnet (Zhang & Shi, 2019)	ICML'19	45.5	56.8	50.5	34.1	36.5	35.3	52.6	78.8	63.1	–	–	–
	APN (Xu et al., 2020)	NeurIPS'20	65.3	69.3	67.2	41.9	34.0	37.6	56.5	78.0	65.5	–	–	–
	DAZLE (Huynh & Elhamifar, 2020a)	CVPR'20	56.7	59.6	58.1	**52.3**	24.3	33.2	60.3	75.7	67.1	–	–	–
	CN (Skorokhodov & Elhoseiny, 2021)	ICLR'21	49.9	50.7	50.3	44.7	**41.6**	43.1	60.2	77.1	67.6	–	–	–
	TransZero (Chen et al., 2022a)	AAAI'22	**69.3**	68.3	**68.8**	52.6	33.4	40.8	61.3	82.3	70.2	–	–	–
	MSDN (Chen et al., 2022c)	CVPR'22	**68.7**	67.5	**68.1**	52.2	34.2	41.3	62.0	74.5	67.7	–	–	–
	I2DFormer (Naeem et al., 2022)	NeurIPS'22	35.3	57.6	43.8	–	–	–	**66.8**	76.8	71.5	35.8	**91.9**	51.5
♠	f-VAEGAN (Xian et al., 2019b)	CVPR'18	48.7	58.0	52.9	45.1	38.0	41.3	57.6	70.6	63.5	56.8	74.9	64.6
	TF-VAEGAN (Narayan et al., 2020)	CVPR'19	53.7	61.9	57.5	48.5	37.2	42.1	58.7	76.1	66.3	62.5	84.1	71.7
	LsrGAN (Vyas et al., 2020)	ECCV'20	48.1	59.1	53.0	44.8	37.7	40.9	54.6	74.6	63.0	–	–	–
	AGZSL (Chou et al., 2021)	ICLR'21	48.3	58.9	53.1	29.9	40.2	34.3	**65.1**	78.9	71.3	–	–	–
	FREE (Chen et al., 2021a)	ICCV'21	54.9	60.8	57.7	47.4	37.2	41.7	60.4	75.4	67.1	**67.4**	84.5	**75.0**
	GCM-CF (Yue et al., 2021)	CVPR'21	61.0	59.7	60.3	47.9	37.8	42.2	60.4	75.1	67.0	–	–	–
	HSVA (Chen et al., 2021b)	NeurIPS'21	52.7	58.3	55.3	48.6	39.0	**43.3**	59.3	76.6	66.8	–	–	–
	ICCE (Kong)	CVPR'22	67.3	65.5	66.4	–	–	–	65.3	82.3	**72.8**	66.1	86.5	74.9
	FREE+ESZSL (Çetin et al., 2022)	ICLR'22	51.6	60.4	55.7	48.2	36.5	41.5	51.3	78.0	61.8	65.6	82.2	72.9
	TF-VAEGAN+ESZSL (Çetin et al., 2022)	ICLR'22	51.1	63.3	56.6	44.0	39.7	41.7	55.2	74.7	63.5	63.5	83.2	72.1
	f-VAEGAN (Xian et al., 2019b) + **DSP**	Ours	62.5	**73.1**	67.4	**57.7**	**41.3**	**48.1**	63.7	**88.8**	**74.2**	**66.2**	**86.9**	**75.2**

Experiments

Comparison with SOTA

Table 2. Comparison with generative ZSL methods on the CUB, SUN, and AWA2 datasets under CZSL setting.

Methods	**CUB**	**SUN**	**AWA2**
	acc	acc	acc
CLSWGAN (Xian et al., 2018)	57.3	60.8	68.2
f-VAEGAN (Xian et al., 2019b)	61.0	64.7	71.1
CADA-VAE (Schönfeld et al., 2019)	59.8	61.7	63.0
Composer (Narayan et al., 2020)	**69.4**	62.6	71.5
FREE (Chen et al., 2021a)	64.8	65.0	68.9
HSVA (Chen et al., 2021b)	62.8	63.8	70.6
f-VAEGAN + DSP	62.8	**68.6**	**71.6**

Experiments

Ablation Study

Table 3. Ablation study on loss terms, smooth evolvement, and feature enhancement of our DSP. The baseline is f-VAEGAN.

Configurations	CUB			SUN		
	U	S	H	U	S	H
baseline	48.7	58.0	52.9	45.1	38.0	41.3
baseline+DSP (w/o $\mathcal{L}_{Scyc}$)	60.0	63.9	61.9	57.4	38.4	46.0
baseline+DSP (w/o $\mathcal{L}_{S2S}$)	61.9	69.6	65.5	58.1	37.2	45.4
baseline+DSP (w/o $\mathcal{L}_{V2S}$)	58.8	61.0	59.9	55.1	34.8	42.7
baseline+DSP (w/o smooth evolvement)	54.4	55.3	54.8	50.3	36.2	42.1
baseline+DSP (w/o enhancement)	52.4	53.9	53.1	54.2	35.0	42.5
baseline+DSP (full)	62.5	73.1	67.4	57.7	41.3	48.1

Experiments

Generative ZSL methods with DSP

Table 4. Evaluation of DSP with multiple popular generative ZSL models on three benchmark datasets. Each row pair shows the effect of adding DSP to a particular generative ZSL model.

Generative ZSL Methods	CUB			SUN			AWA2		
	U	S	H	U	S	H	U	S	H
CLSWGAN (Xian et al., 2018)	43.7	57.7	49.7	42.6	36.6	39.4	57.9	61.4	59.6
CLSWGAN (Xian et al., 2018)+**DSP**	**51.4**	**63.8**	**56.9**$^{\uparrow 7.2}$	**48.3**	**43.0**	**45.5**$^{\uparrow 6.1}$	**60.0**	**86.0**	**70.7**$^{\uparrow 11.1}$
f-VAEGAN (Xian et al., 2019b)	48.7	58.0	52.9	45.1	38.0	41.3	57.6	70.6	63.5
f-VAEGAN (Xian et al., 2019b)+**DSP**	**62.5**	**73.1**	**67.4**$^{\uparrow 14.5}$	**57.7**	**41.3**	**48.1**$^{\uparrow 6.8}$	**63.7**	**88.8**	**74.2**$^{\uparrow 10.7}$
TF-VAEGAN (Narayan et al., 2020)	53.7	61.9	57.5	48.5	37.2	42.1	58.7	76.1	66.3
TF-VAEGAN (Narayan et al., 2020)+**DSP**	**58.7**	**67.4**	**62.8**$^{\uparrow 5.3}$	**60.3**	**45.3**	**51.7**$^{\uparrow 9.6}$	**65.6**	**87.1**	**74.8**$^{\uparrow 8.5}$
FREE (Chen et al., 2021a)	54.9	60.8	57.7	47.4	37.2	41.7	60.4	75.4	67.1
FREE (Chen et al., 2021a)+**DSP**	**60.9**	**68.7**	**64.6**$^{\uparrow 6.9}$	**60.3**	**44.1**	**51.0**$^{\uparrow 9.3}$	**65.3**	**89.2**	**75.4**$^{\uparrow 8.3}$

Experiments

Qualitative Evaluation

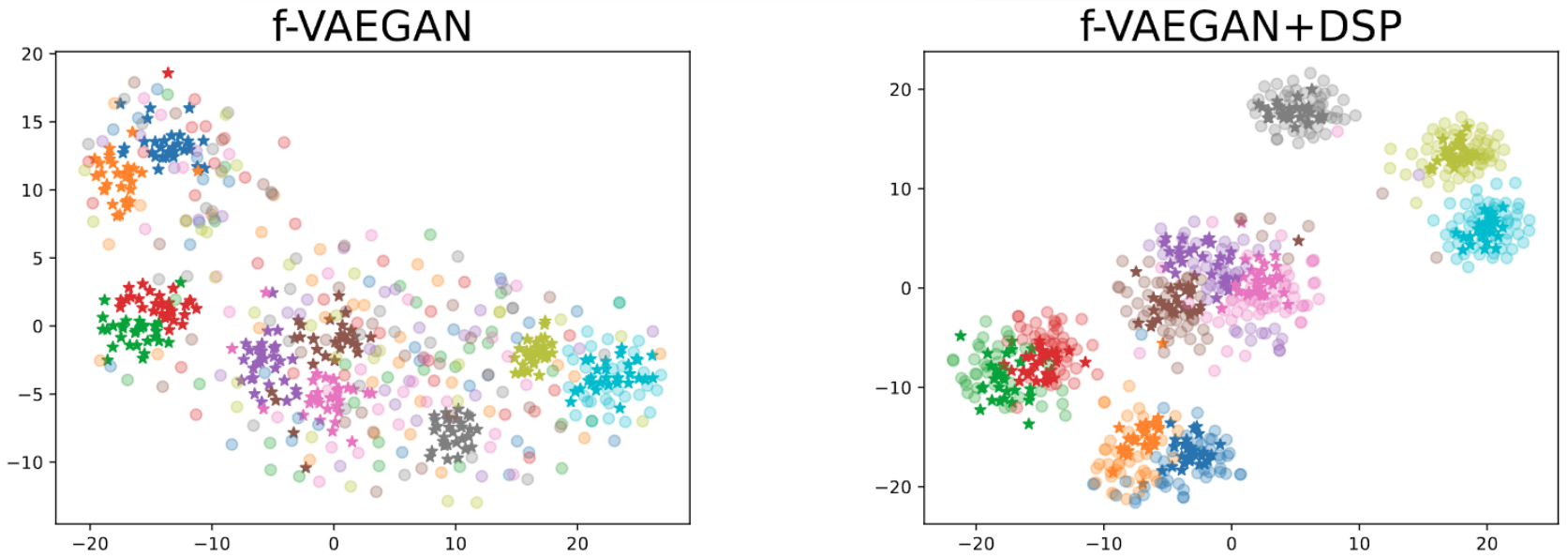
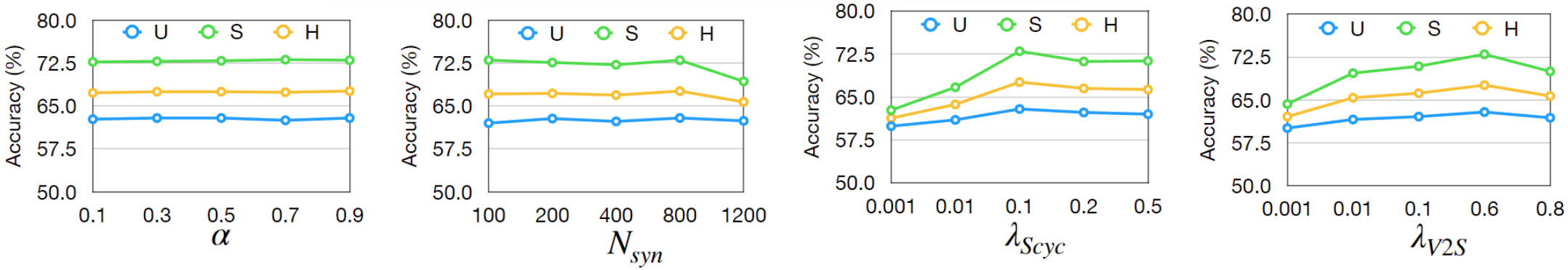


Figure 3. Qualitative evaluation with t-SNE visualization. The sample features from f-VAEGAN are shown on the left, and from f-VAEGAN with our DSP integration are shown on the right. We use 10 colors to denote randomly selected 10 classes from CUB. The ○ and ⋆ are denoted as the real and synthesized sample features, respectively. The synthesized sample features and the real features distribute differently on the left while distributing similarly on the right. (Best Viewed in Color)

Experiments

Hyper-parameter Analysis



(a) Varying α effect (b) Varying N_{syn} effect (c) Varying λ_{Scyc} effect (d) Varying λ_{V2S} effect

Figure 4. Hyper-parameter analysis. We show the GZSL performance variations on CUB by adjusting the value of α in (a), the value of N_{syn} in (b), the value of loss weight λ_{Scyc} in (c), and the value of loss weight λ_{V2S} in (d). (Best Viewed in Color)