

Generalized Zero-shot Learning with Multi-source Semantic Embeddings for Scene Recognition

Xinhang Song^{1,2}, Haitao Zeng¹, Sixian Zhang^{1,2}, Luis Herranz³, Shuqiang Jiang^{1,2}

¹Institute of Computing Technology, Chinese Academy of Sciences

²University of Chinese Academy of Sciences

³Computer Vision Center, Barcelona, Spain

MM20-fp2464



中国科学院计算技术研究所
Institute of Computing Technology, Chinese Academy of Sciences



Outline

- Introduction
- Framework & Proposed method
- Experiments
- Conclusion



Outline

- Introduction
- Framework & Proposed method
- Experiments
- Conclusion

■ Background

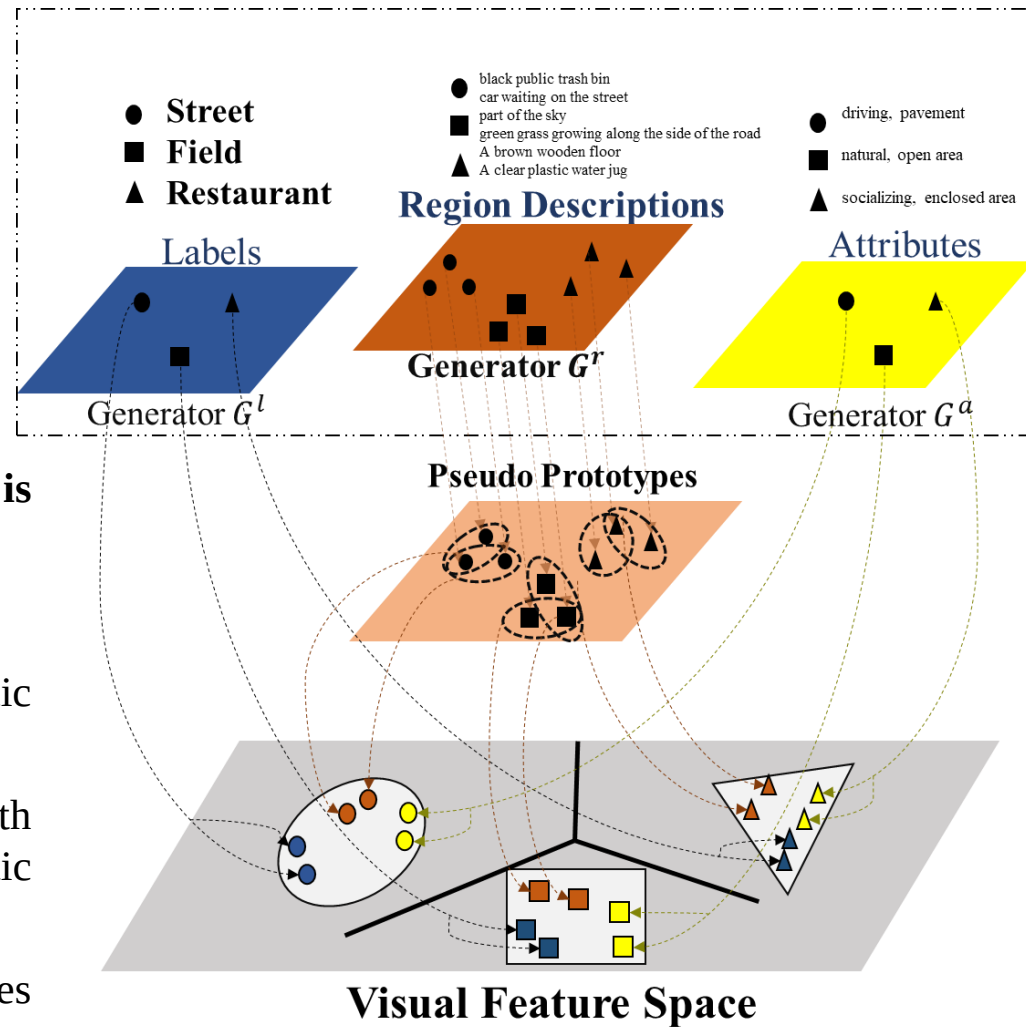
- Previous ZSL works usually focus on objects
- Objects compose the scenes
- One scene may contain many prototypes

■ Motivation

- **ZSL for scene recognition is typical challenging**

■ Contributions

- **A dataset** with multi-semantic annotation of scenes
- **A framework** of GZSL with multi-source semantic embedding
- **Two-step** scene prototypes learning

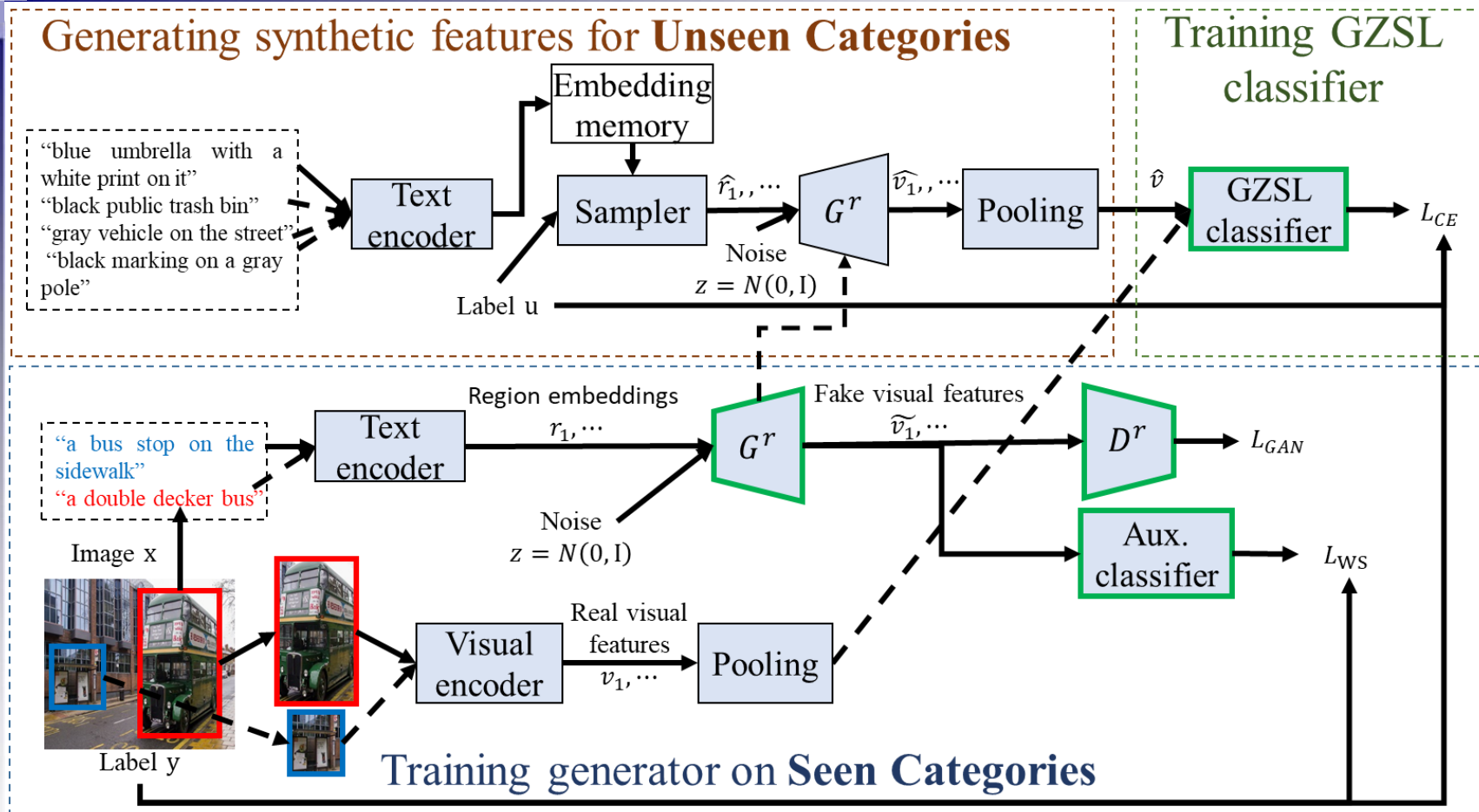




Outline

- Introduction
- Framework & Proposed method
- Experiments
- Conclusion

Framework

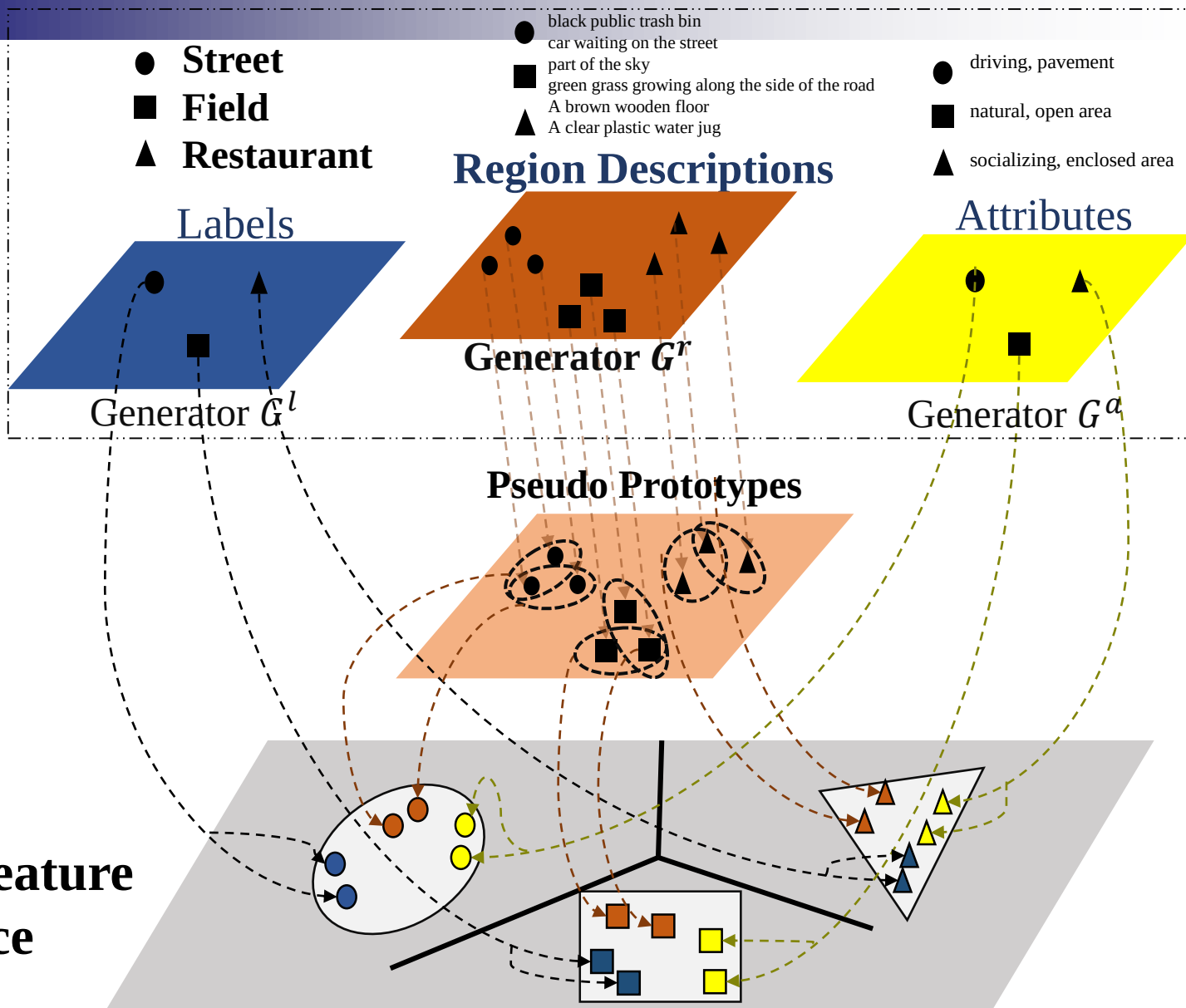


- 1) training generator with seen categories (bottom line);
- 2) generating synthetic features for unseen categories;
- 3) training GZSL joint classifier of seen and unseen categories.



Proposed

Multi-source Semantic Embedding





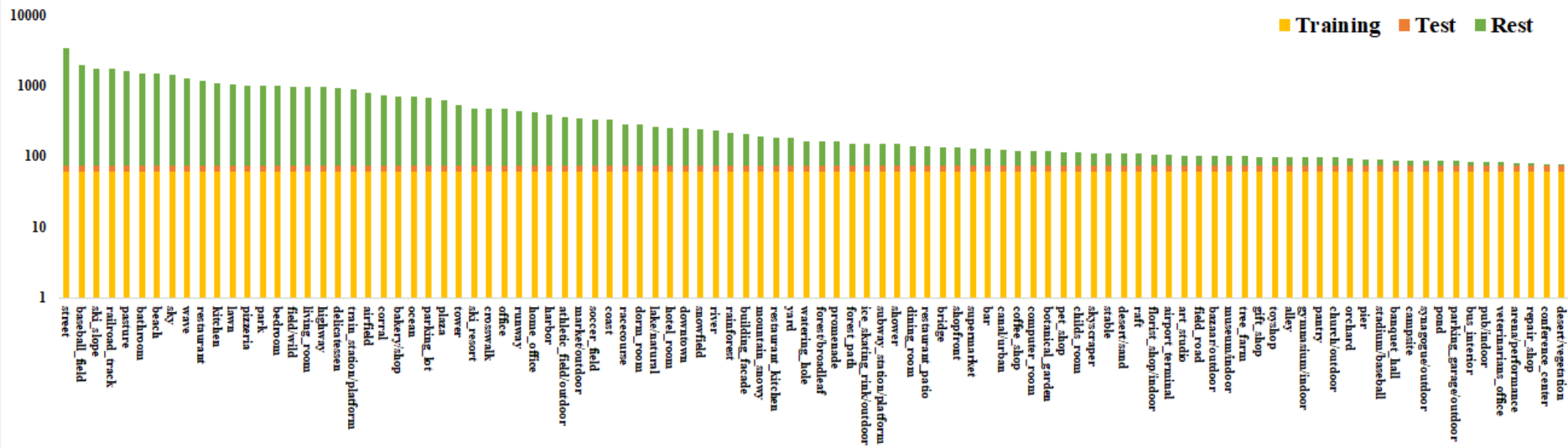
Outline

- Introduction
- Framework & Proposed method
- Experiments
- Conclusion



Experimental setting

Image Amount Distribution (Seen)



ZSSR Dataset

- Categories
 - Seen:100, Unseen:100
- Annotations: Labels, Attributes, Descriptions (Global and Local)
- #Images
 - Seen:44K, (60 per category), Unseen: 5K, (15 per category)



Experimental results

#Unseen	SAE			SLE			SDE		
	U	S	H	U	S	H	U	S	H
100	4.53	61.60	8.45	3.80	62.0	7.16	2.07	60.73	4.00
300	14.40	48.07	22.16	12.27	51.13	19.79	6.00	51.67	10.75
500	19.01	38.44	25.44	15.64	40.39	22.55	8.98	40.40	14.70
1000	22.58	31.32	26.24	19.03	31.66	23.77	10.90	22.05	14.59
2000	17.64	24.93	20.66	15.79	27.60	20.08	9.80	21.42	13.45



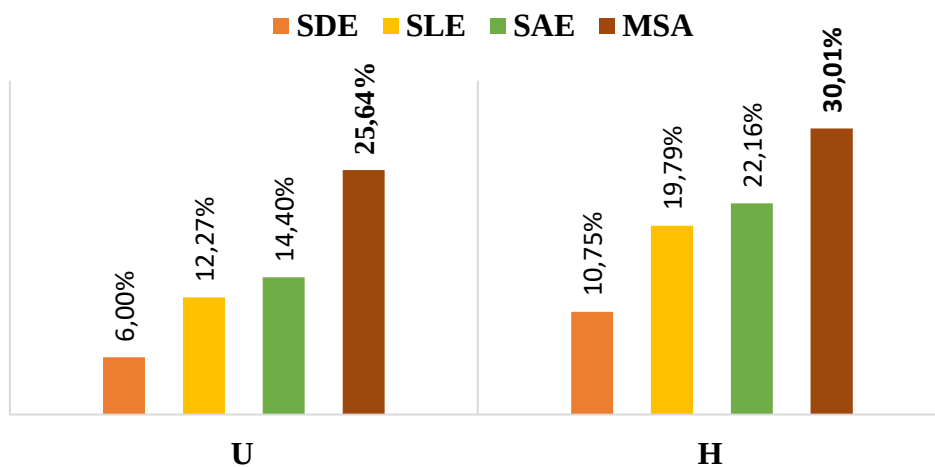
More generated unseen features lead to better results.

- Scene attribute embedding (SAE)
- Scene label embedding (SLE)
- Scene description embedding (SDE)
- Region description embedding (RDE)

Semantic aggregation leads to large improvement.



GZSL with Different Semantic Embedding





Experimental results

Region feature analysis on ZSSR

#Seen	Unseen #syn																	
	S						U						H					
	20	40	60	80	100	200	20	40	60	80	100	200	20	40	60	80	100	200
10	47.87	31.13	28.67	26.87	26.40	26.67	12.13	22.07	23.40	22.20	20.13	16.00	19.36	25.83	25.77	24.31	22.84	20.00
20	59.07	37.80	36.33	36.40	34.00	28.07	4.60	19.47	20.07	20.93	20.73	20.93	8.54	25.70	25.85	26.58	25.76	23.98
30	61.46	44.00	38.47	39.07	39.27	32.13	3.67	16.73	19.73	19.87	20.06	21.93	6.92	24.25	26.09	26.34	26.56	26.07
40	61.20	52.00	38.27	37.00	38.27	36.27	2.60	14.40	19.00	18.87	19.60	20.80	4.99	22.55	25.39	24.99	25.92	26.44
50	64.13	51.73	39.80	39.27	42.80	37.47	2.07	13.73	18.13	19.27	18.27	20.53	4.00	21.70	24.92	25.85	25.61	26.53
60	63.93	52.67	43.87	39.33	38.33	39.73	1.47	11.93	17.00	18.73	19.33	19.60	2.87	19.46	24.50	25.38	25.70	26.25

Comparison to SUN Attribute

	Seen	Unseen	Accuracy (%)		
			U	S	H
SAE	SUN	SUN	46.70	39.75	42.94
	ZSSR	SUN	60.00	35.25	44.10
	SUN+ZSSR	SUN	61.25	50.75	55.51
SLE	SUN	SUN	27.50	39.75	32.51
	ZSSR	SUN	38.00	32.75	35.18
	SUN+ZSSR	SUN	36.25	49.50	41.85
RDE (ours)	ZSSR	SUN	33.65	40.50	36.76
	SUN+ZSSR	SUN	36.55	47.00	41.12
MSA (SSR) (ours)	ZSSR	SUN	61.75	45.75	52.56
	SUN+ZSSR	SUN	60.75	59.50	60.12
Related Works					
LisGAN (CVPR19)	SUN	SUN	49.20	44.33	46.64
CAVA (ICCV19)	SUN	SUN	49.65	49.47	49.56
ABP (CVPR19)	SUN	SUN	49.45	43.11	46.06

- ↓
- Regions features outperform Global features
 - For region features: not the more generated the better

- ⇒
- Our ZSSR can improve the performances on SUN Attribute
 - Our ZSSR obtains the state-of-the-art on SUN Attribute



Outline

- Introduction
- Framework & Proposed method
- Experiments
- Conclusion



Conclusion

- Contributions:
 - **A dataset** with multi-semantic annotation
 - **A framework** of GZSL with multi-source semantic embedding
 - **Two-step** scene prototypes learning
- Experiments
 - Regions features work better than global features
 - Multi-source embedding works



中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

Thanks for Watching!