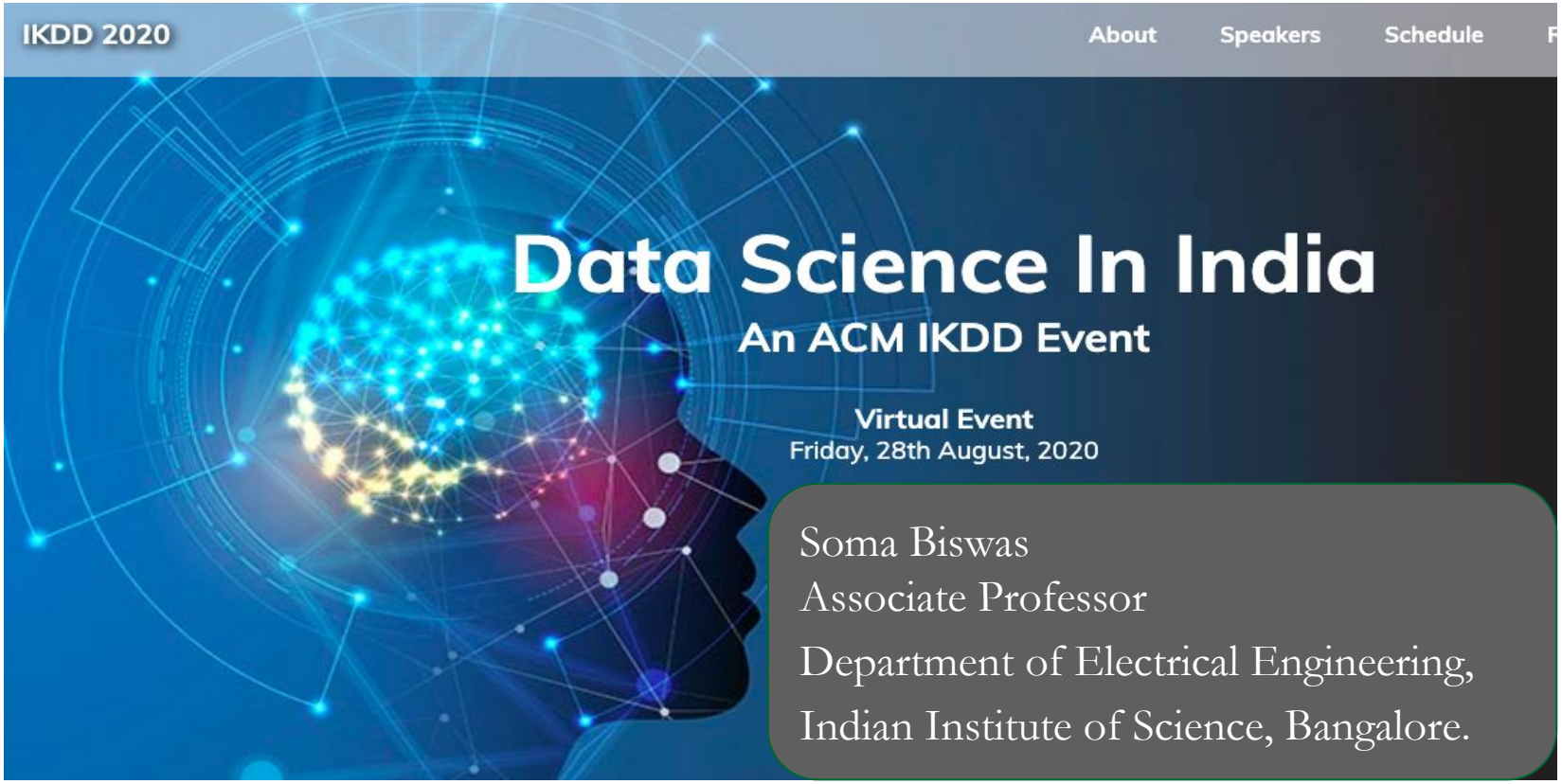
Sketch Based Image Retrieval



Soma Biswas
Associate Professor
Department of Electrical Engineering,
Indian Institute of Science, Bangalore.

Computer Vision - Why is it important?

- Vision is useful: Images and video are everywhere!

Difficult for Computers!

- To extract "meaning" from pixels

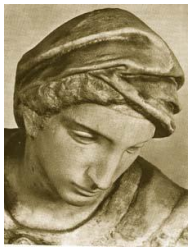
0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

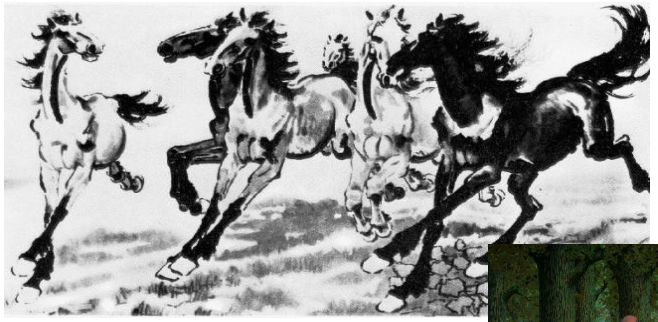
What we see

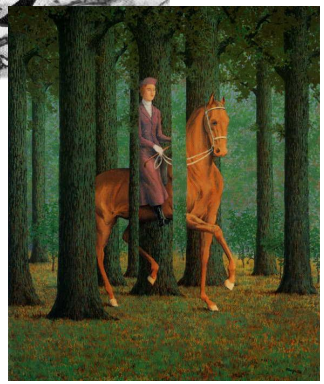
What a computer sees

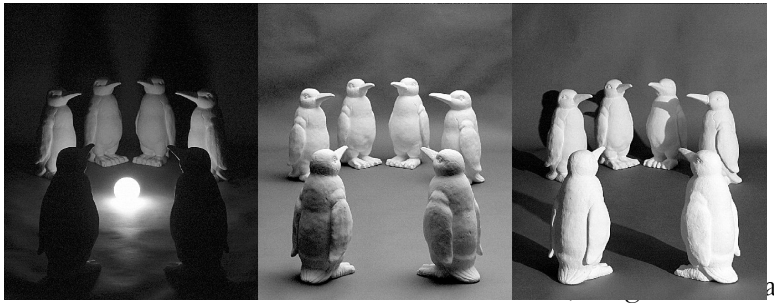
Source: S. Narasimhan

Challenges in Computer Vision



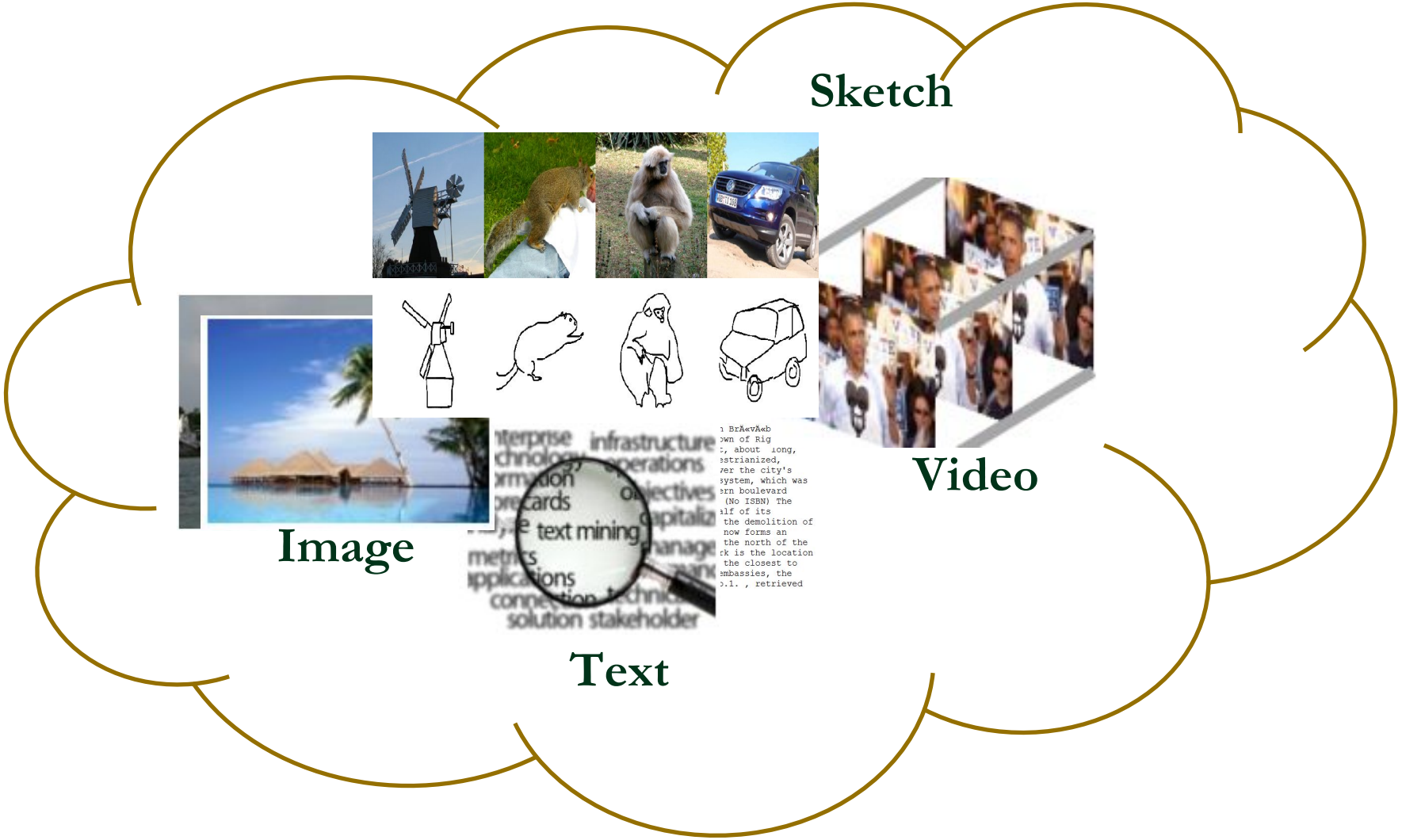




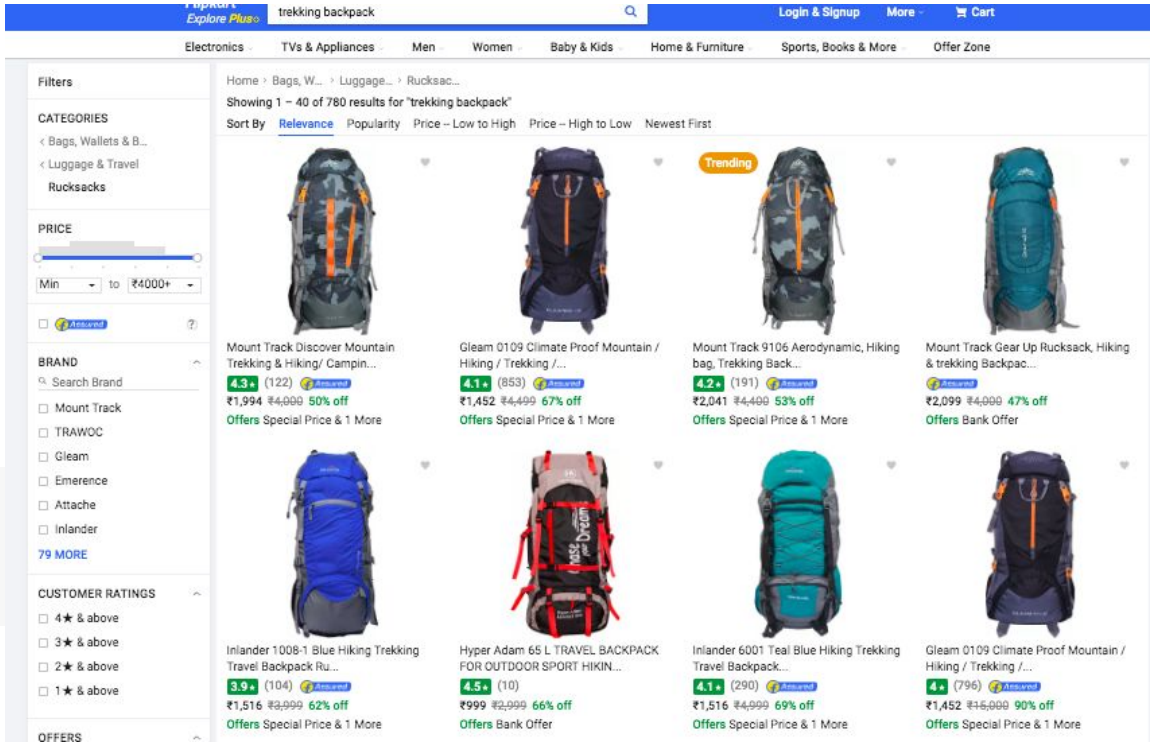


alba

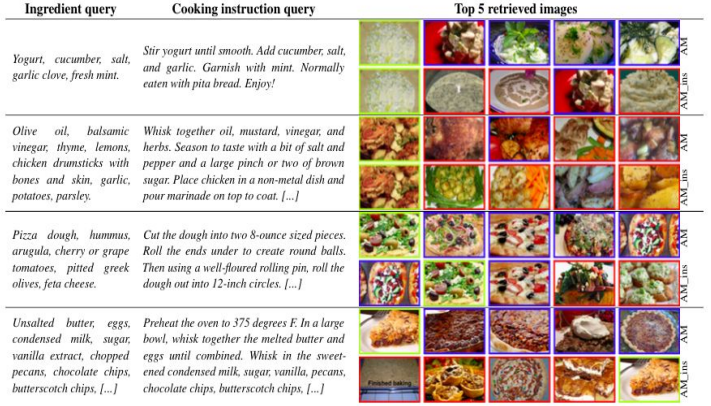
But data stored in different domains/modalities!



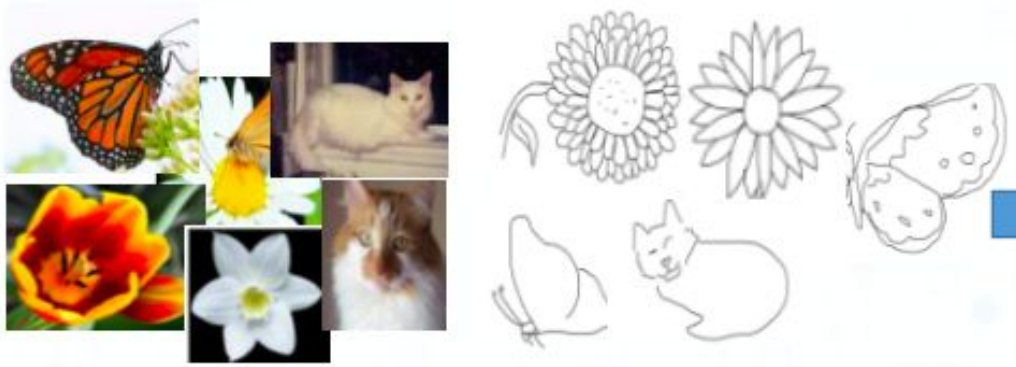
Cross-Modal Retrieval - SBIR



Applications

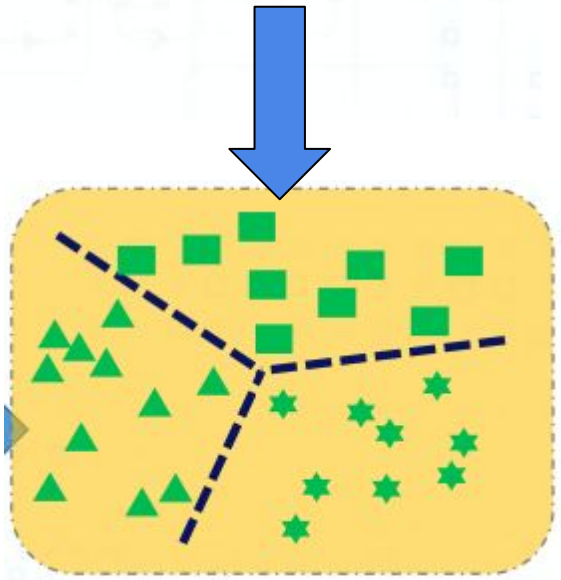


Overall Idea of SBIR



Training Images and Sketches
From category : {Cat, Flower, Butterfly}

Learn Cross-domain matching model



Latent-space Embedding

Challenges

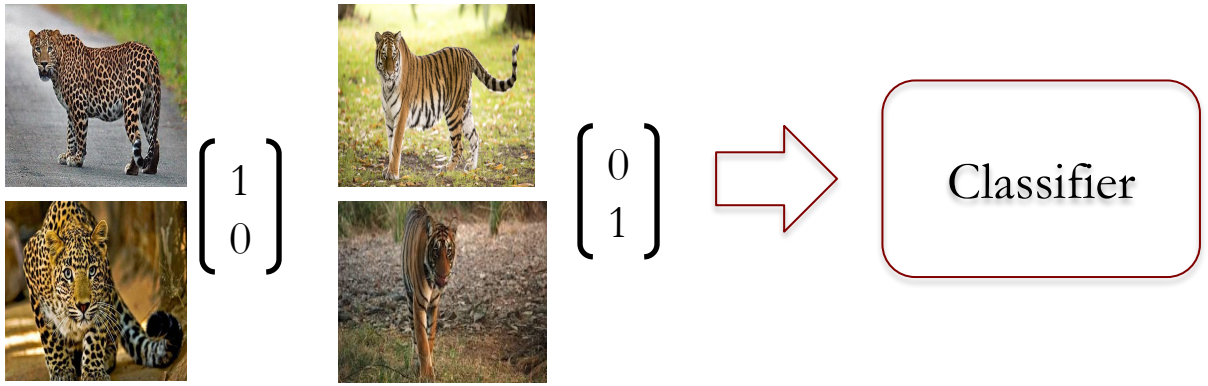
1.Significant differences between the sketches and images.
2. Significant variations within the same classes of both sketches and images.

Zero-Shot ZBIR

- The number of categories is dynamically increasing.
- We have seen all examples during training is a strong assumption.
- What if during testing, the query sketch comes from an unseen/novel class, which is not seen during training.

Linking the Seen & Unseen Classes

- **Conventional Supervised Classification**



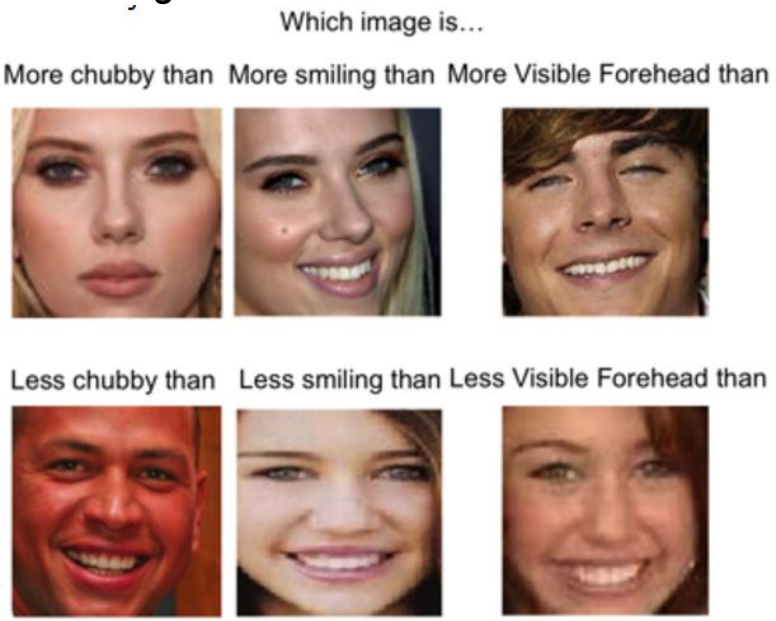
- **Zero-Shot Learning**: Given an image which belongs to a novel category, predict the class.

- Attributes of the categories are mapped with the features.
- Attributes are **shared** across categories, both seen and unseen.

Attributes

- Attributes can be *Manual*, *Relative* or *Class-name word-embeddings*.

(a) Manual Attributes



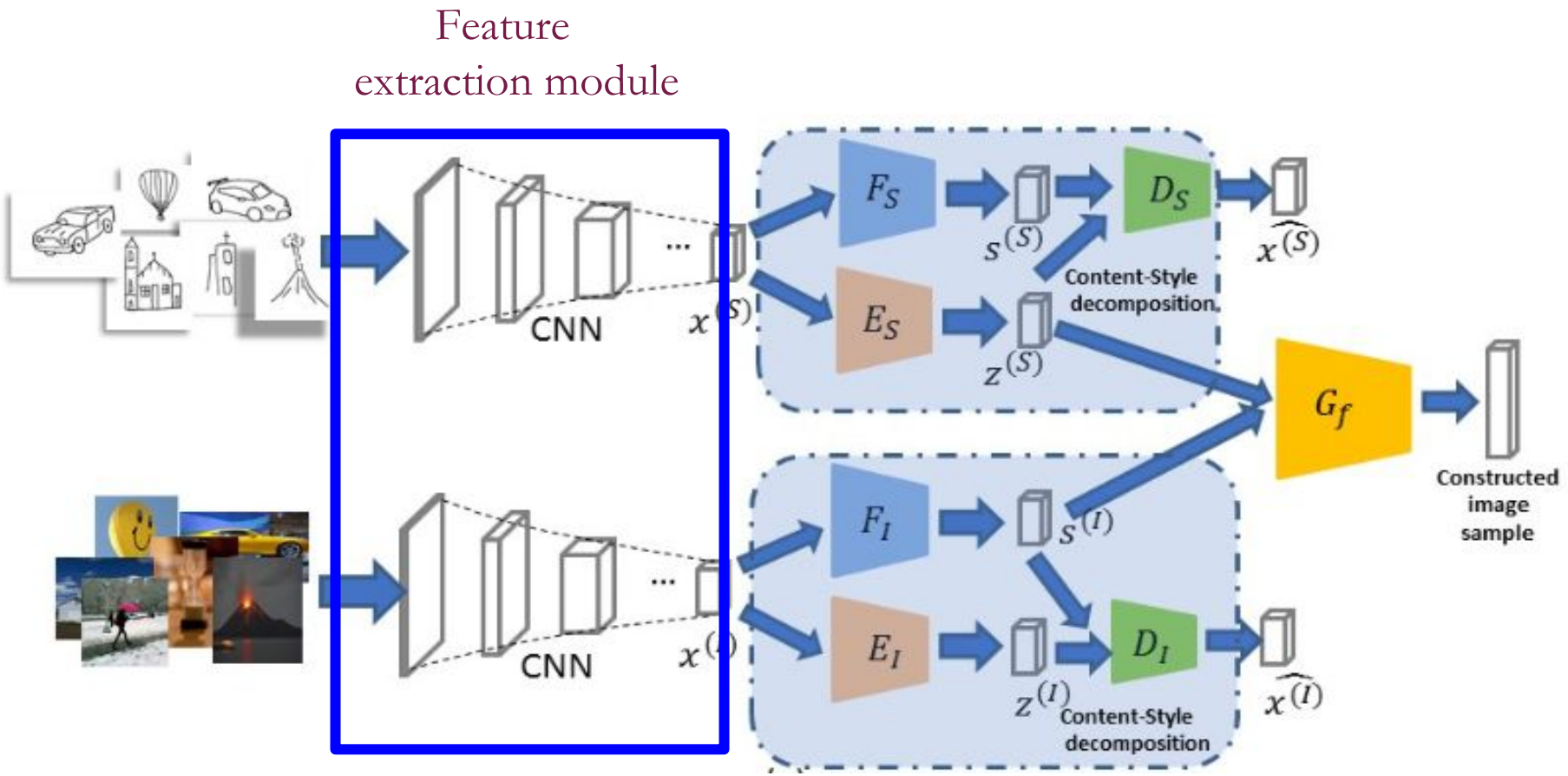
(b) Relative Attributes

** (a) Images and corresponding attributes are taken from *Animals with Attributes*(AwA) dataset, proposed by Lampert et al., CVPR, 2009[1].
** (b) Images and relative attribute annotations are from *Relative Face Attribute* dataset, proposed by Biswas et al., CVPR 2013[2]

[1] C. H. Lampert, H. Nickisch and S. Harmeling, Learning to detect unseen object classes by between-class attribute transfer, *CVPR, 2009*.
[2] A. Biswas and D. Parikh, Simultaneous active learning of classifiers and attributes via relative feedback, *CVPR, 2013*.

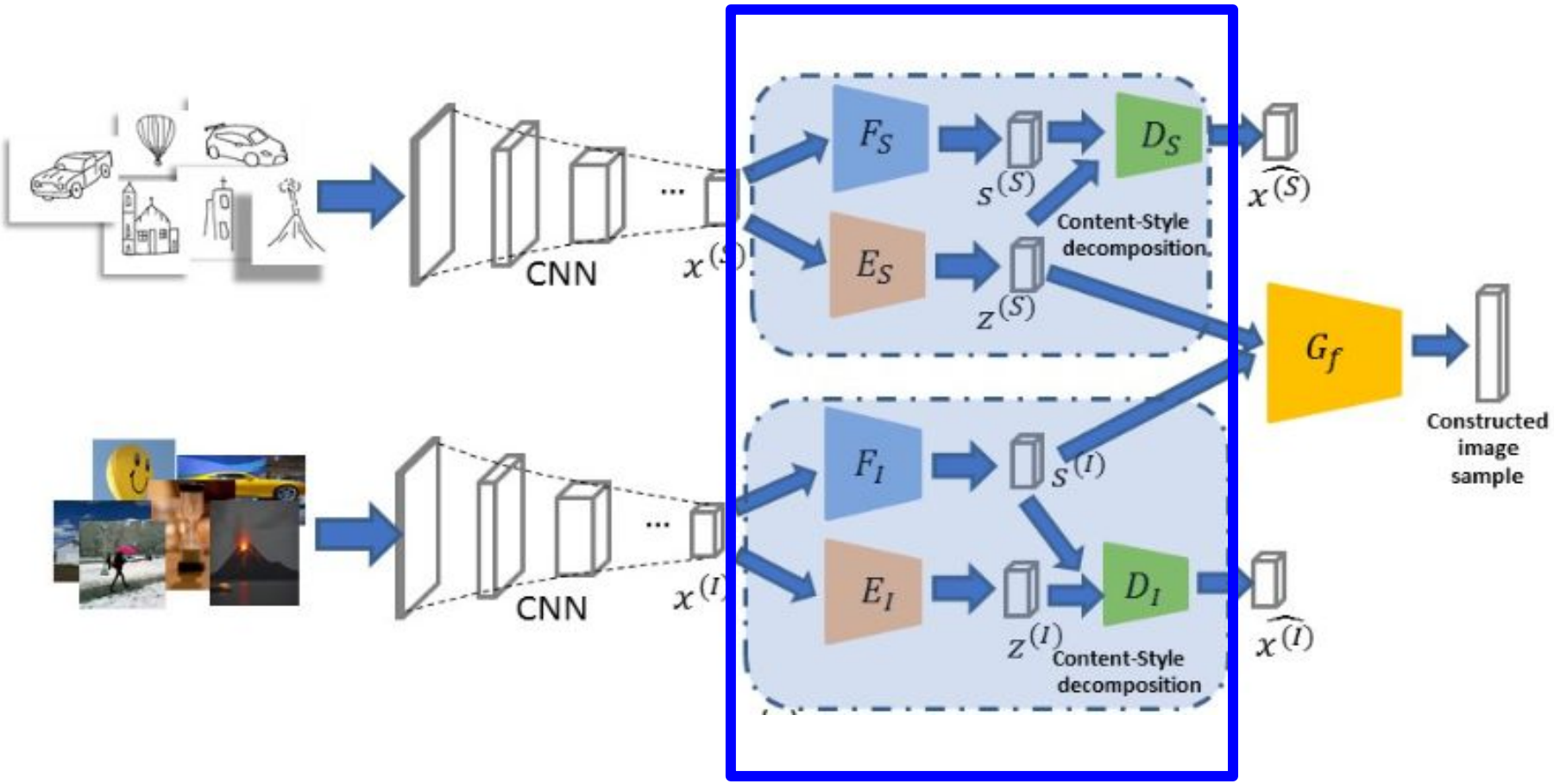
- Word-embedding representations (GloVe, Word2Vec) of class-names can be used as Attributes.

ZS-SBIR - Training



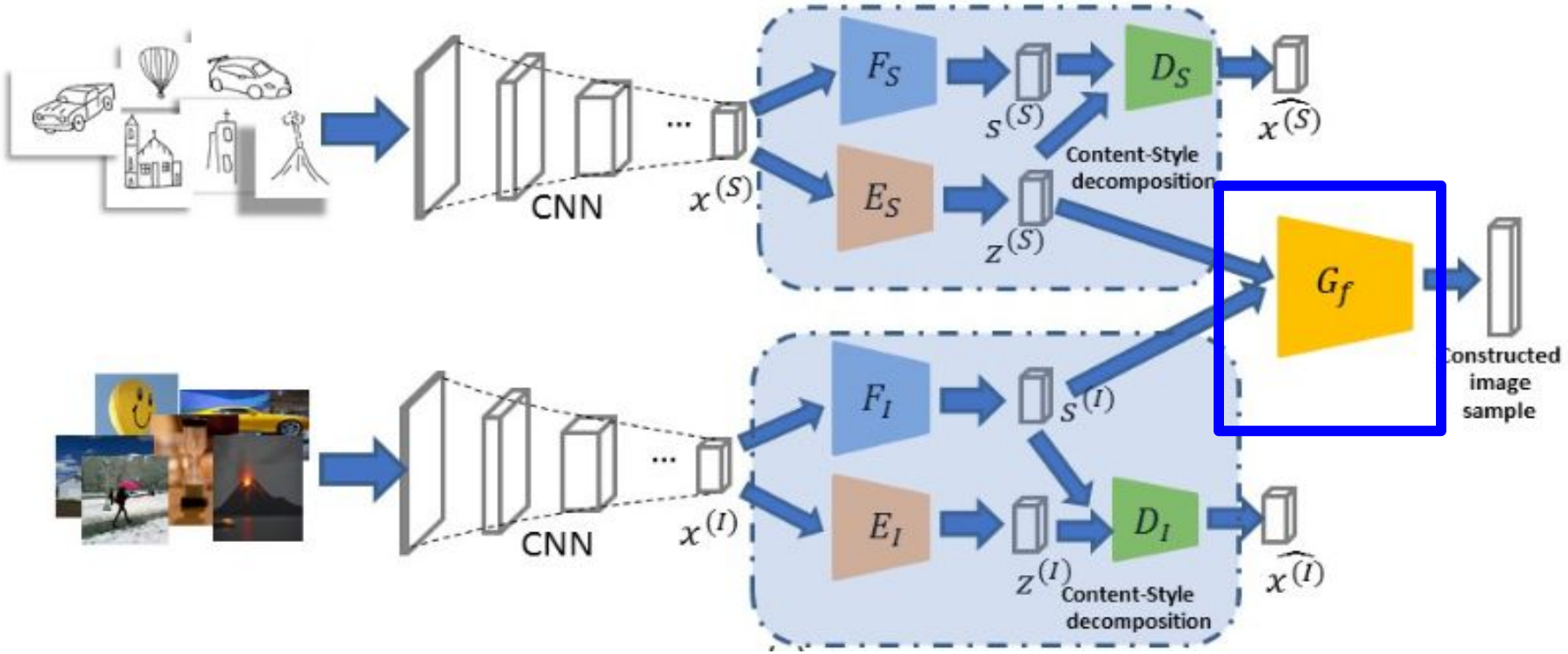
ZS-SBIR - Training

Content-Style
decomposition module



ZS-SBIR - Training

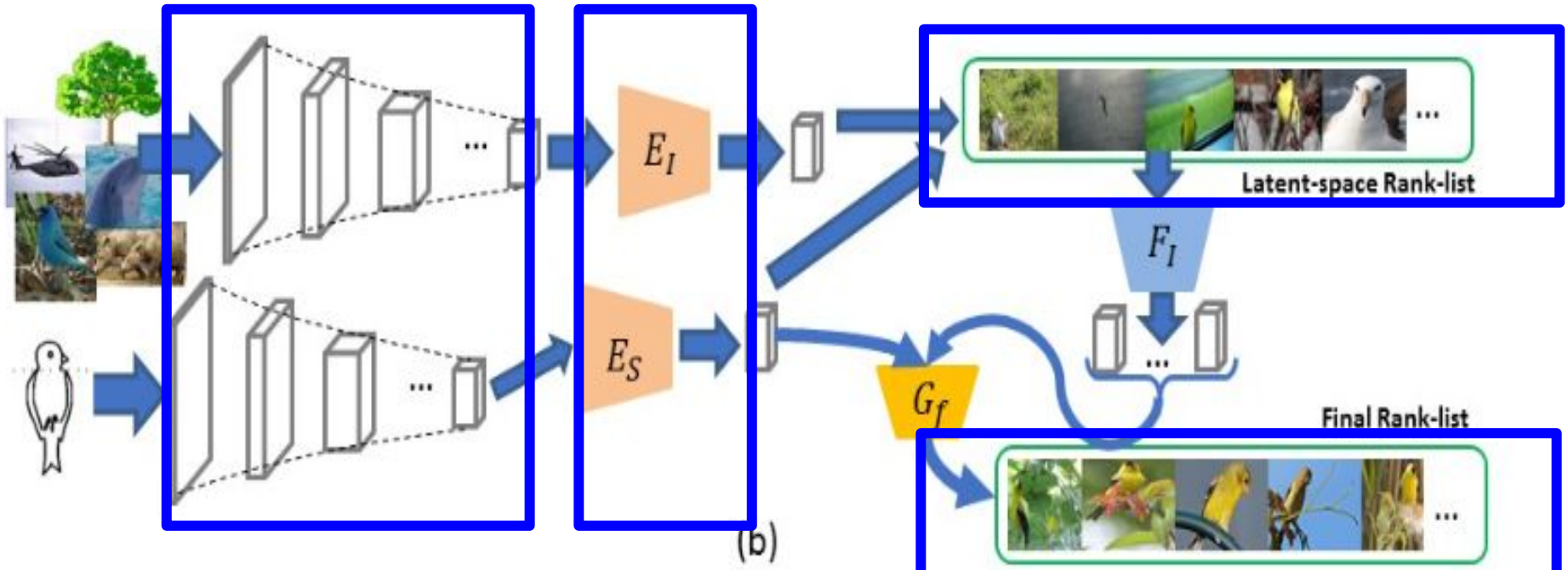
Content-Style fusion module



Retrieval

Content-Style decomposition module

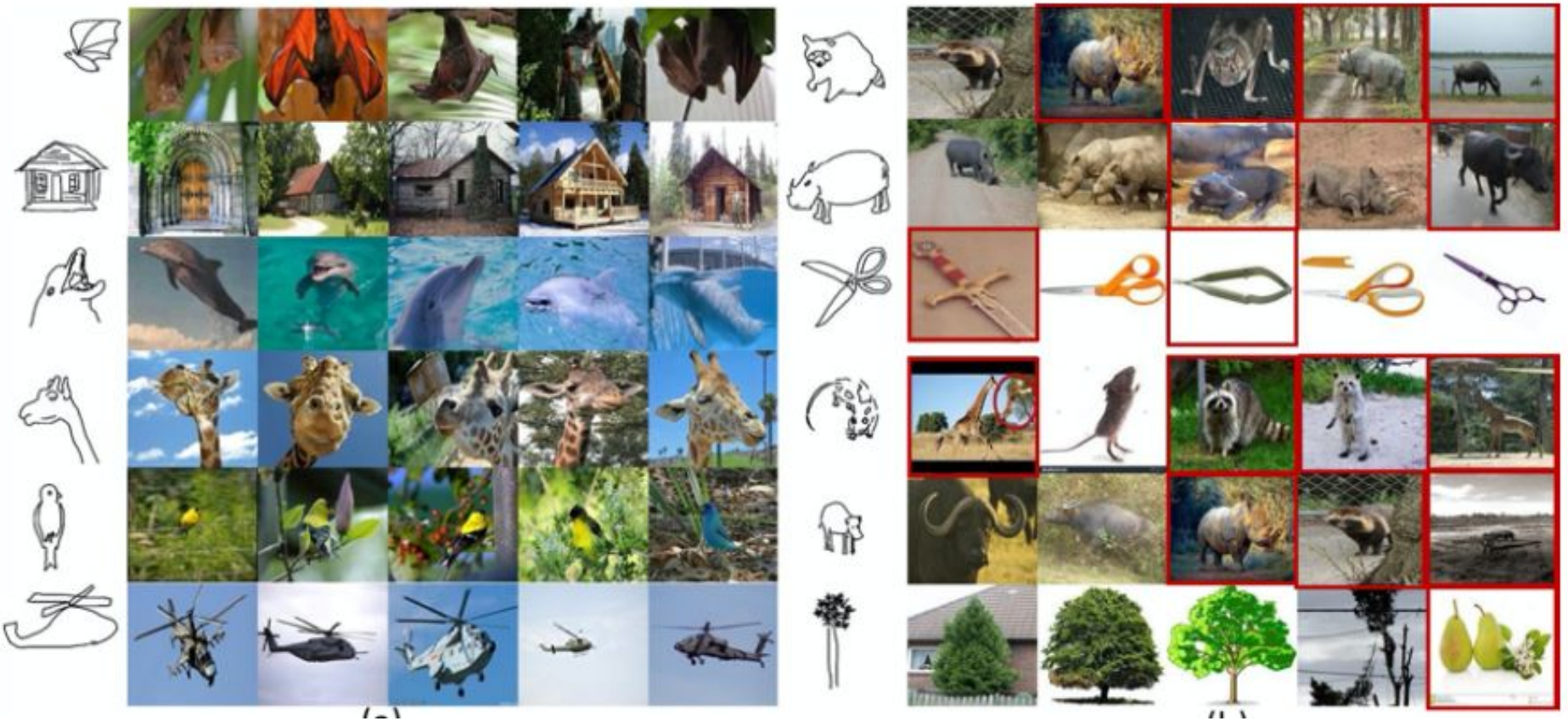
Latent-space Matching



Feature extraction module

Image-space Matching

Qualitative Results

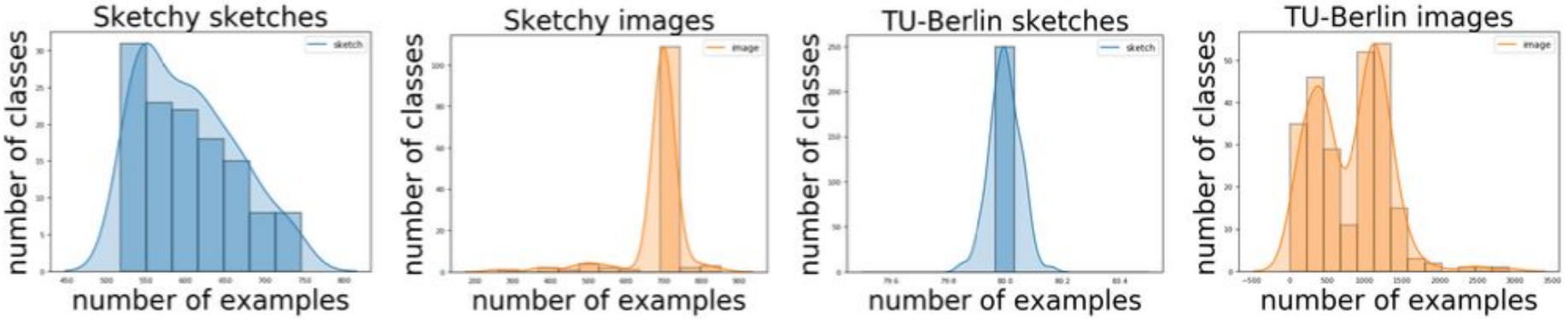


Quantitative Results

Type	Evaluation methods	Precision@200	MAP@200
SBIR methods	Baseline	0.106	0.054
	Siamese-1 [11]	0.243	0.134
	Siamese-2 [3]	0.251	0.149
	Coarse-grained Triplet [25]	0.169	0.083
	Fine-grained Triplet	0.155	0.081
	DSH [17]	0.153	0.059
ZSL methods	Direct Regression	0.066	0.022
	ESZSL [22]	0.187	0.117
	SAE [14]	0.238	0.136
ZS-SBIR	CAAE [35]	0.260	0.156
	CVAE [35]	0.333	0.225
	Proposed	**0.4001**	**0.3581**

Effect of Imbalance in Training Data

- State-of-the-art ZS-SBIR approaches underperform if training data is imbalanced.



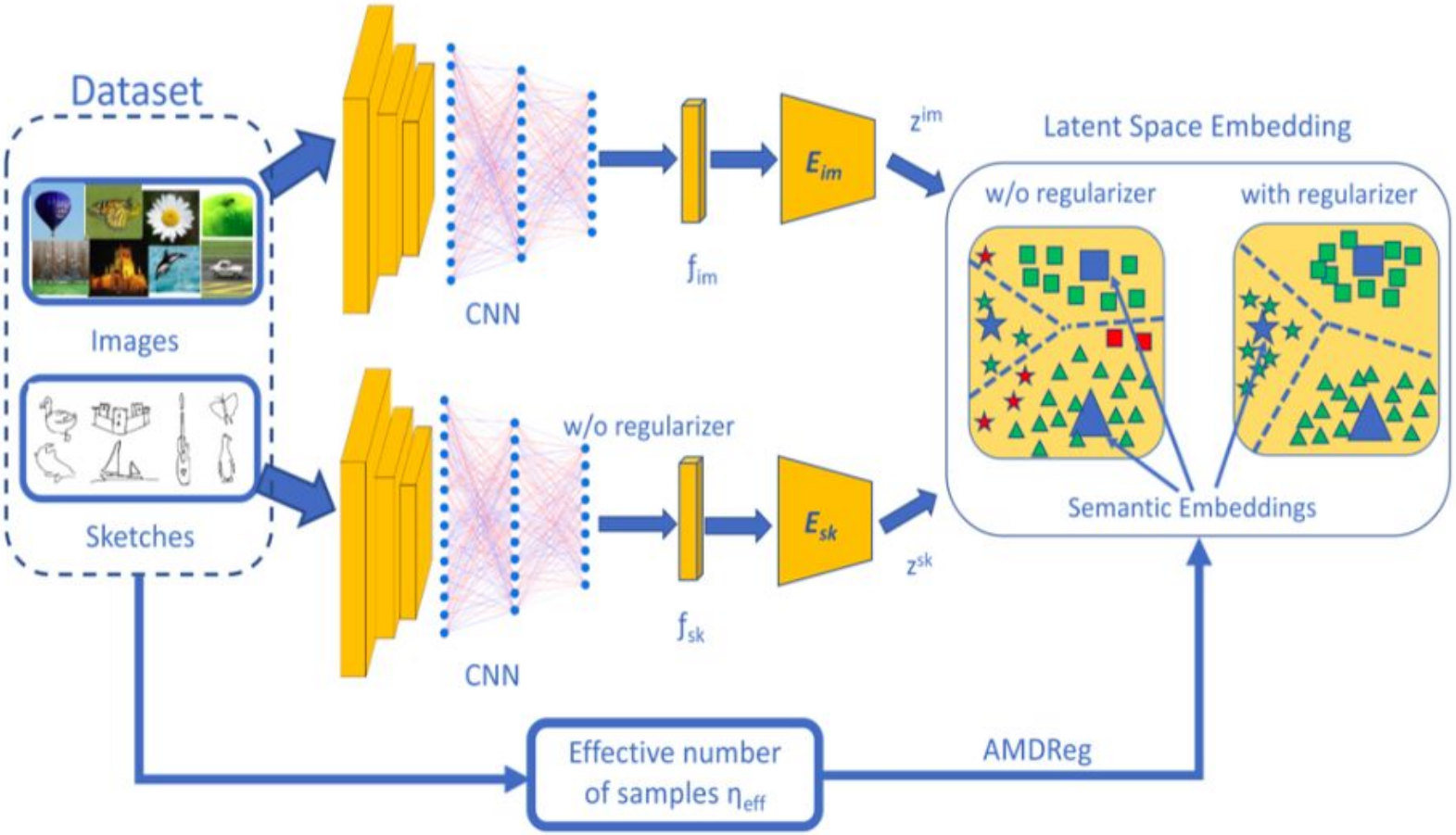
Dataset	Method	Balanced Data	Imbalanced data			
			Long-tailed		Step	
			P=10	P=100	P=10	P=100
mini-Sketchy*	Style-guided network [1]	0.395	0.234	0.185	0.241	0.156

Table 1. Performance (MAP@200) of state-of-the-art ZS-SBIR models for imbalanced data

* a balanced sub-set of Sketchy-extended dataset with 60-classes, each having 500 images and sketches per class.

[1] T. Dutta and S. Biswas, "Style-guided zero-shot sketch-based image retrieval," *BMVC*, 2019

Handling Imbalance in ZS-SBIR



Proposed AMDReg

We propose Adaptive-margin Diversity Regularizer as additional loss component with any existing ZS-SBIR model during training.

$$R(P) = \frac{1}{C_{seen}} \sum_{i<j} \{||p_i - p_j||_2^2 - (d_{mean} + \Delta_j)\}^2,,$$
$$\forall j \in \{1,2,\dots,C_{seen}\}$$

$$\Delta_j = \frac{K}{n^j_{effective}}$$

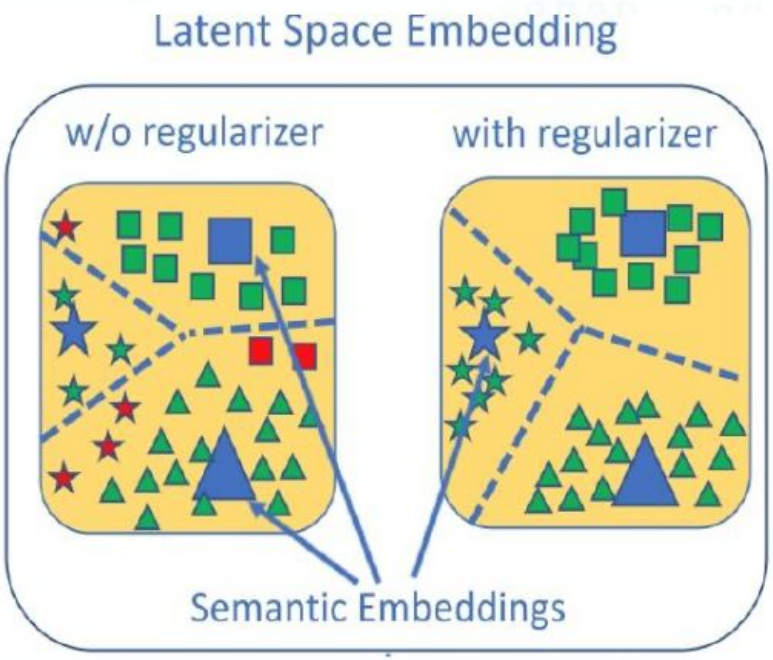


Figure 2. Illustration of proposed AMDReg in learning the latent-space embeddings

ZS-SBIR Approaches with AMDReg

Method	Sketchy-extended		TU-Berlin	
	MAP@all	Precision@100	MAP@all	Precision@100
Style-guided network [1]	0.375	0.484	0.254	0.355
Style-guided network [1] + AMDReg	**0.410**	**0.512**	**0.291**	**0.376**
SEM-PCYC [2]	0.349	0.463	0.297	0.426
SEM-PCYC [2]+ AMDReg	**0.397**	**0.494**	**0.330**	**0.473**
SAKE [3]	0.547	0.692	0.428*	0.534*
SAKE [3]+ AMDReg	**0.551**	**0.715**	**0.447**	**0.574**

Table 2. Performance of state-of-the-art ZS-SBIR model with proposed AMDReg

* produced using the original code provided by the authors.

[1] T. Dutta and S. Biswas, "Style-guided zero-shot sketch-based image retrieval,'' *BMVC*, 2019
[2] A. Dutta and Z. Akata, ``Semantically-tied paired cycle consistency for zero-shot sketch-based image retrieval," *CVPR*,2019
[3] Q. Liu, L. Xie, H. Wang and A. Yuille, "Semantic-aware knowledge preservation for zero-shot sketch-based image retrieval," *ICCV*, 2019

Qualitative Results

Conclusion & Future Directions

- Cross-modal retrieval is a very important research area.
- With touch screens, etc. SBIR will have several applications in future.
- Generalizing to completely unseen classes without any prior information is important for real-world applications.
- Handling challenges like data imbalances, etc. can positively affect the performance.

- Fusing multiple modalities, like sketch and text as input.
- Accounting for domain differences in the data.
- …...

Thank You