

## Context

### Goal: Text illustration

Textual query: **a man cycling on a mountain**

Retrieve most appropriate images:

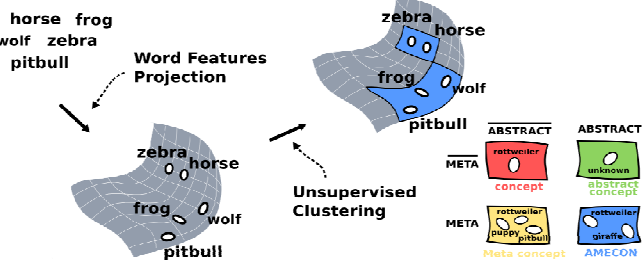


## Contributions

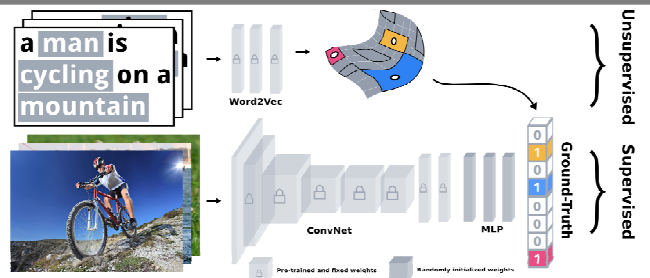
➤ **Mixing supervised and unsupervised learning for Text-Image Matching (TIM) problems**

- 1) Learn a codebook from word2vec features of texts > each cluster: Abstract Meta-Concept (**Text2Amecon**)
- 2) Learn a MLP from CNN features of images to Abstract Meta-Concepts (**Image2Amecon**)
- 3) Solving TIM problems: (i) Text2Amecon (ii) Image2Amecon (iii) **matching in AMECON Space**

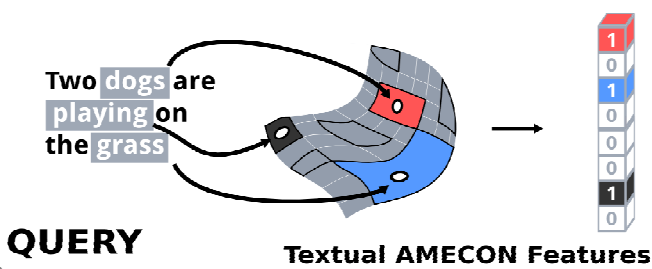
## Learning Textual Amecons



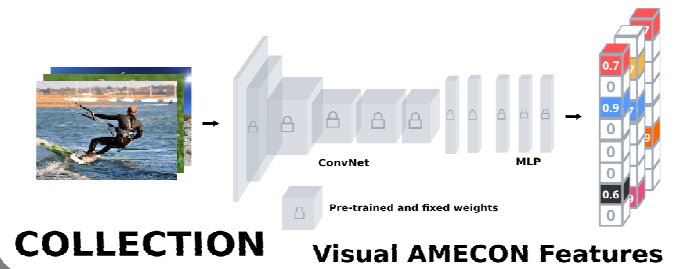
## Learning Visual Amecons



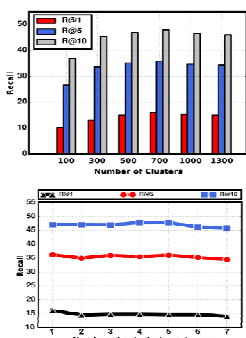
## Text2Amecon



## Image2Amecon



## Experiments



Method	Denotation	Flickr-8k			Flickr-30k		
		R@1	R@5	R@10	R@1	R@5	R@10
Karpathy et al. [17]	DeFrag	9.7	29.6	42.5	10.3	31.4	44.5
Kiros et al. [18]	MNLM	10.4	31.0	43.7	11.8	34.0	46.3
Mao et al. [21]	m-RNN	11.5	31.0	42.4	12.6	31.2	41.5
Karpathy et al. [16]	BRNN*	11.8	32.1	44.7	15.2	37.7	50.5
Yan et al. [36]	DCCA	12.7	31.2	44.1	12.6	31.0	43.0
Tsui et al. [32]	MACC†	10.2	29.3	41.4	12.4	33.5	46.1
Our Approach	AMECON	15.9	37.9	49.5	18.3	41.3	53.5

- [16] Deep Visual-Semantic Alignments for Generating Image Descriptions, CVPR 2015  
 [17] Deep Fragment Embeddings for Bidirectional Image Sentence Mapping, NIPS 2014  
 [18] Unifying Visual-Semantic Embeddings with Multimodal Neural Language Models, TAEL 2015  
 [21] Explain Images with Multimodal Recurrent Neural Networks, ArXiv 2014  
 [32] Aggregating Image and Text Quantized Correlated Components, CVPR 2016  
 [36] Deep Correlation for Matching Images and Text, CVPR 2015

## Matching in AMECON Space

