AMECON: Abstract Meta-Concept Features for Text Illustration

Ines Chami[1,*], Youssef Tamaazousti[2,*] and Hervé Le Borgne[2]
1: Stanford University, USA – 2: CEA LIST, FRANCE
* Both authors contributed equally

Text-illustration System

Textual query: a man cycling on a mountain

Text-illustration system

Most appropriate images:

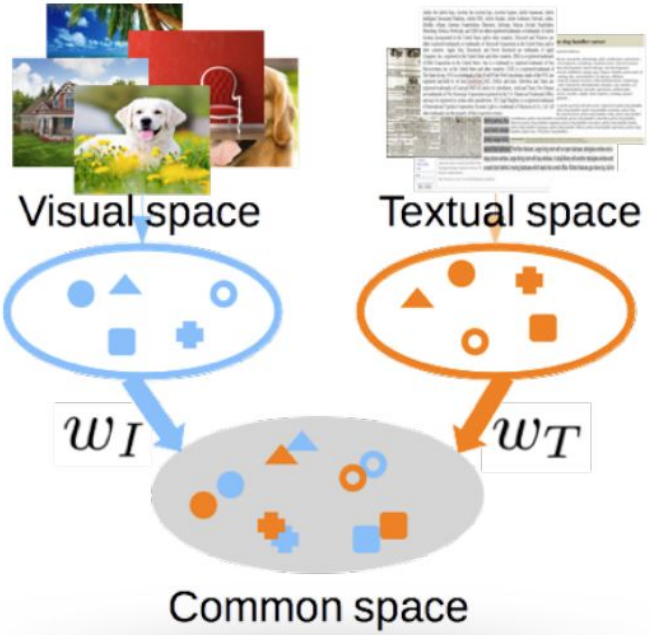



Cross-Modal Retrieval Task

- **Cross-Modal Retrieval task**
  - Given a document in one modality, find (from database) the most relevant documents in another modality
  - Text-illustration
    - Query: sentences
    - Collection: images
- **Hard problem: semantic gap**

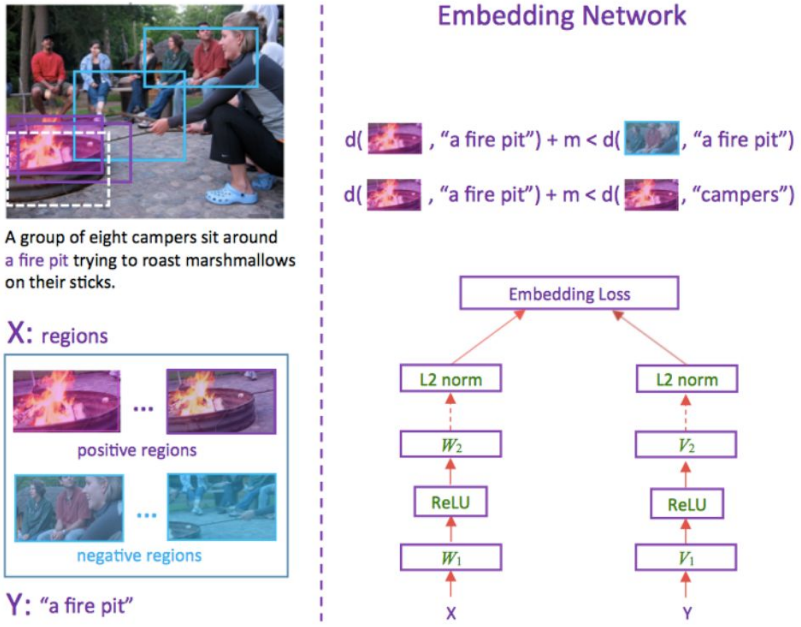
Cross-Modal Retrieval Approach 1

- **Canonical Correlation Analysis**
  - Hardoon et al. Neural Computation 2004
  - Hwang and Grauman, IJCV 2012
  - Costa Pereira et al. TPAMI 2014
  - Tran et al., CVPR 2016
  - etc.



Cross-Modal Retrieval Approach 2

- **Neural Network (NN)**
  - Karpathy and Fei-Fei, NIPS 2014
  - Yan and Mikolajczyk, CVPR 2015
  - Karpathy and Fei-Fei, CVPR 2015
  - Mao et al., ICLR 2015
  - Kiros et al., TACL 2015
  - Wang et al., CVPR 2016
  - etc.



Main Principle of NN Approach

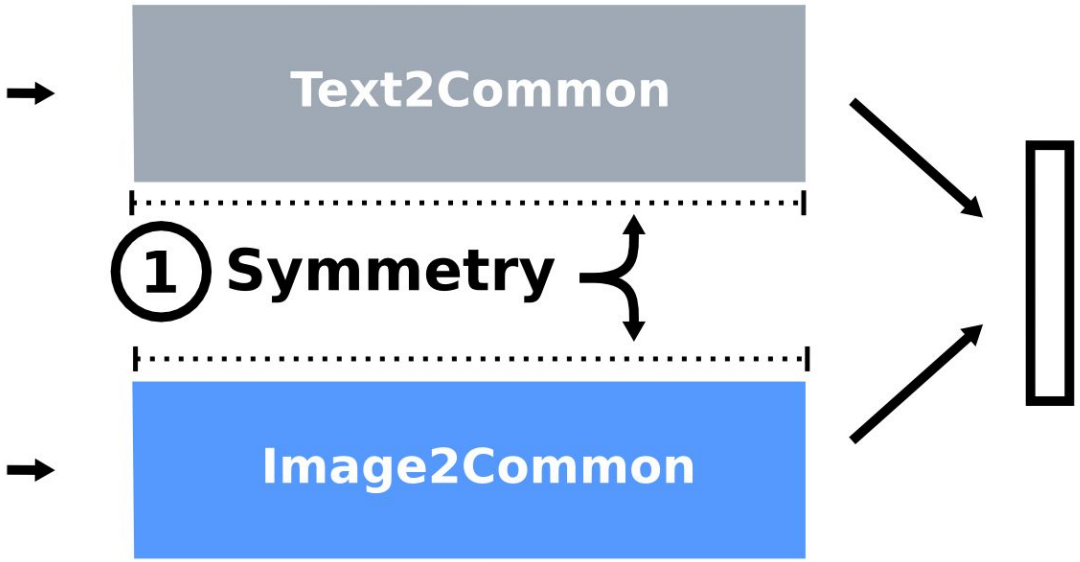
a man is kite-surfing on the water

Text2Common

Image2Common

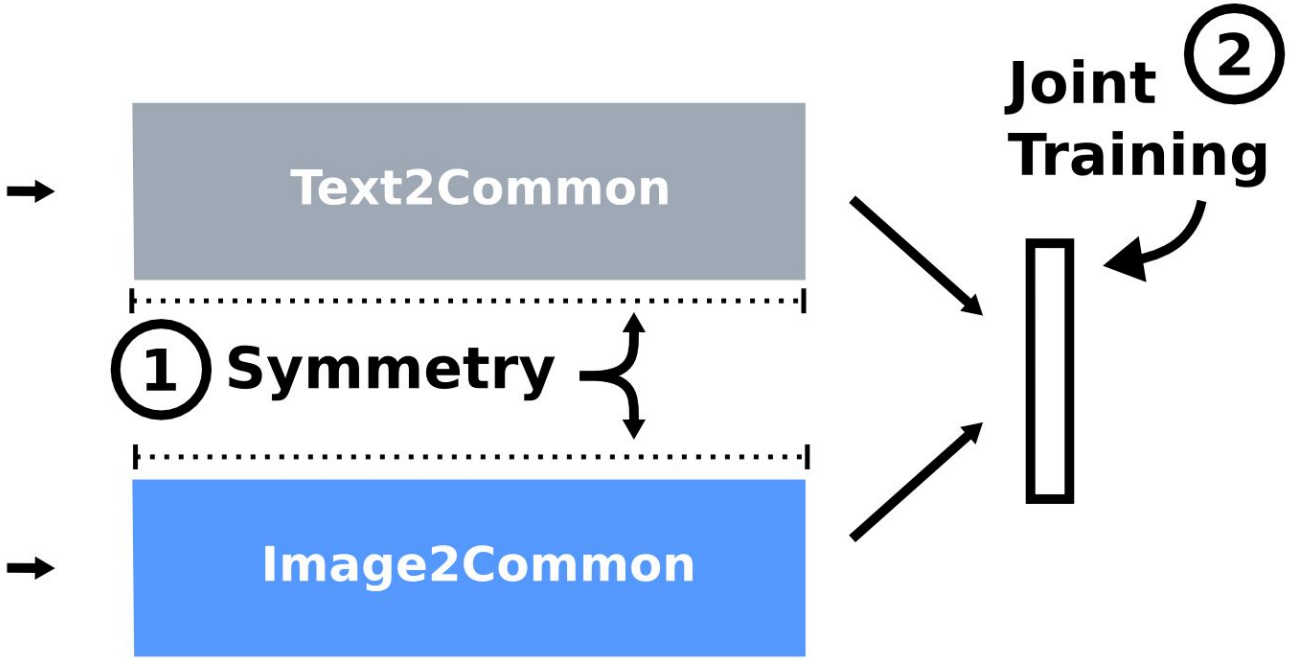
Main Principle of NN Approach

a man is kite-surfing on the water

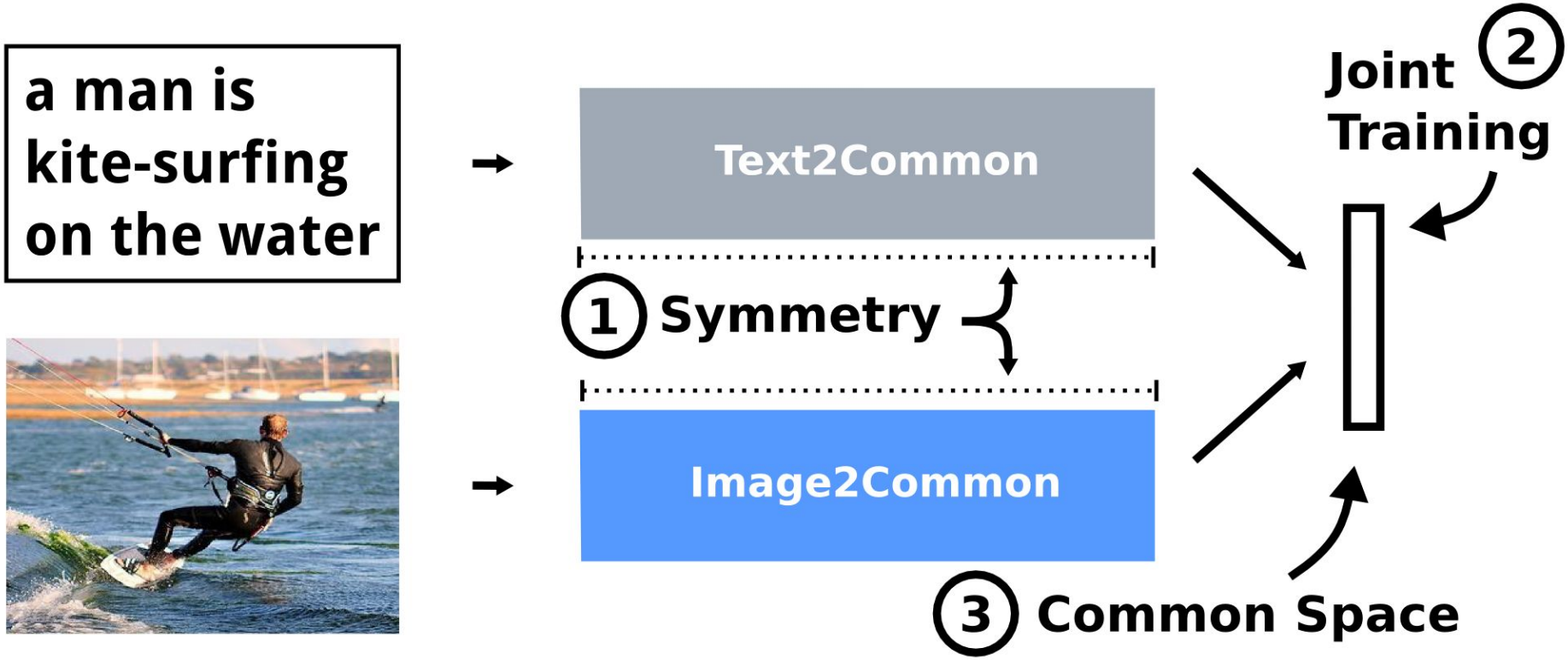


Main Principle of NN Approach

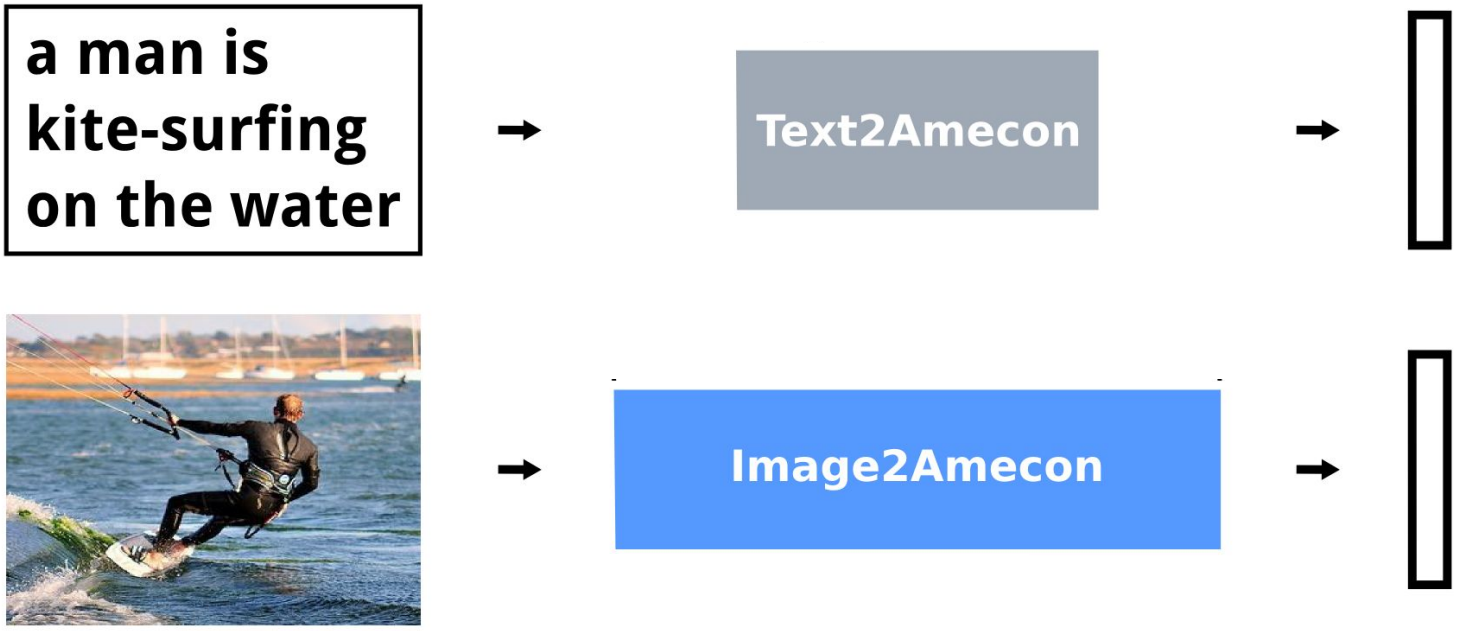
a man is kite-surfing on the water



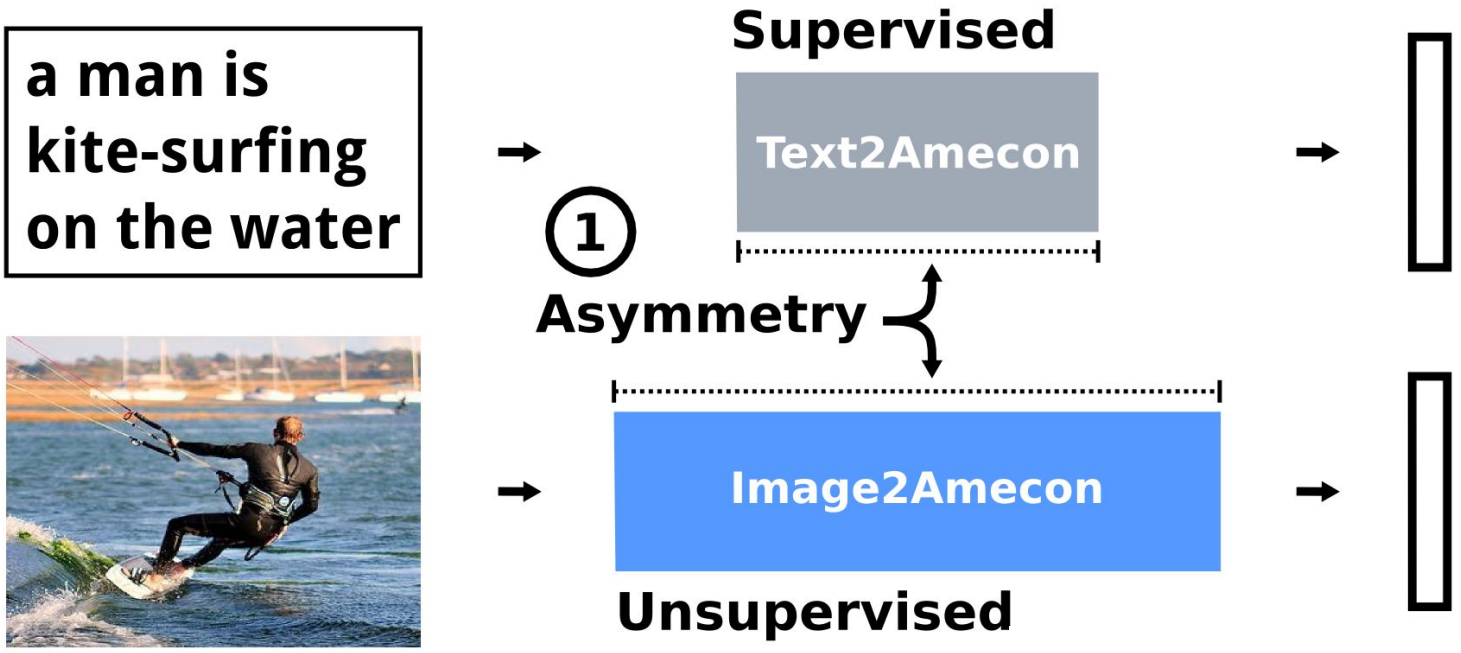
Main Principle of NN Approach



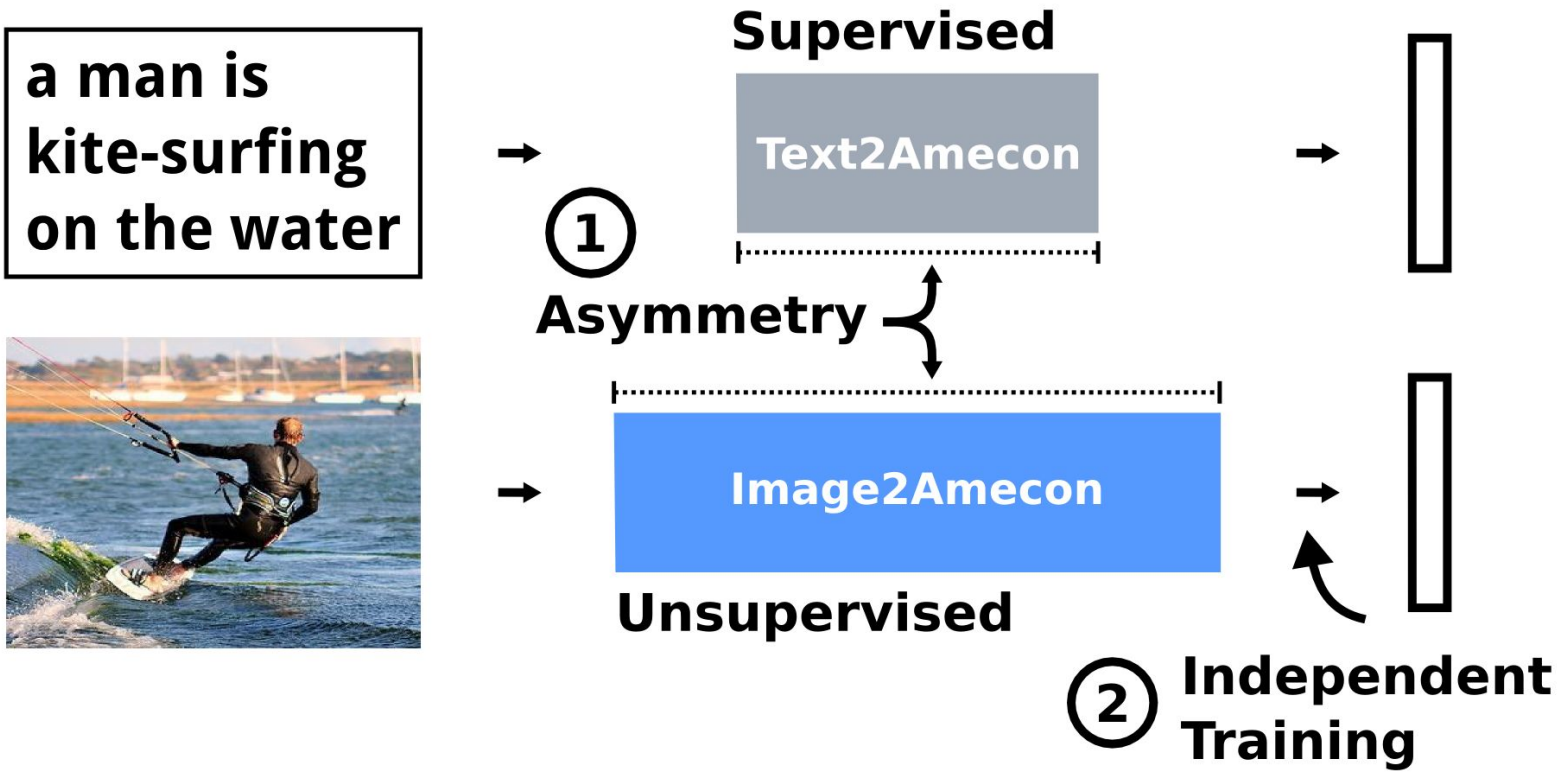
This work: New Approach



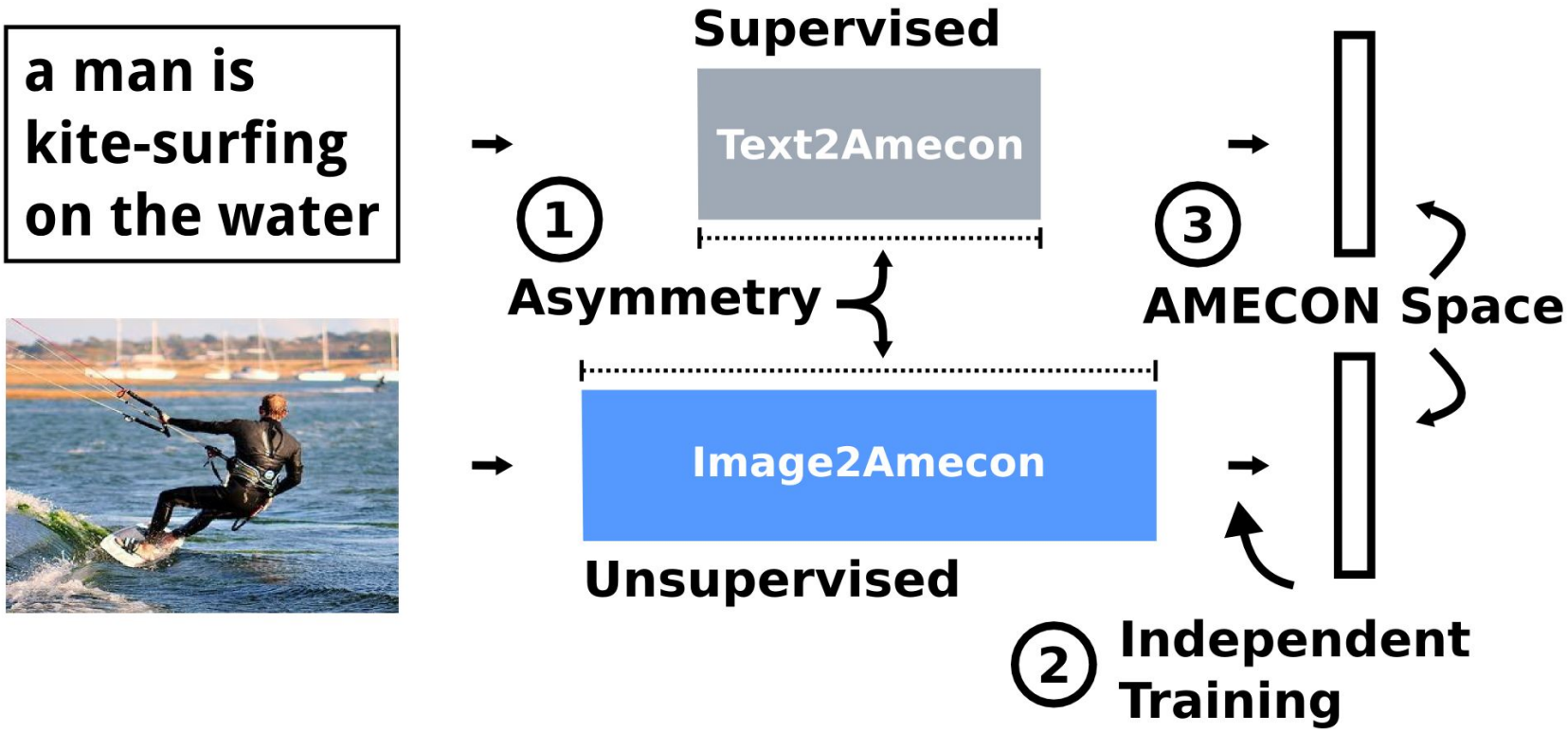
This work: New Approach



This work: New Approach



This work: New Approach

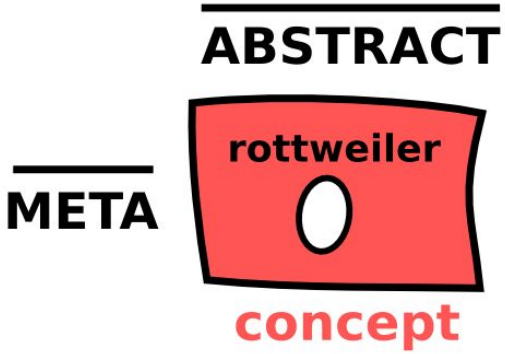


AMECON principle

- **AMECON: Abstract Meta-CONcept**
  - Abstract-concept + Meta-concept

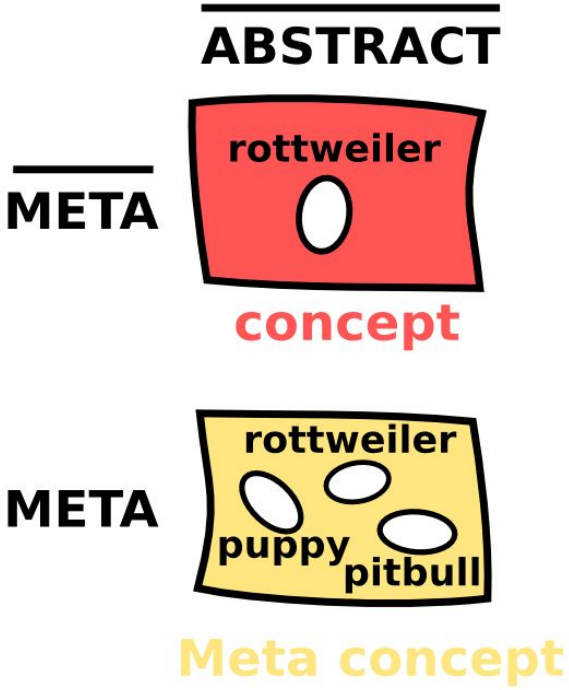
AMECON principle

- **AMECON: Abstract Meta-CONcept**
  - Abstract-concept + Meta-concept



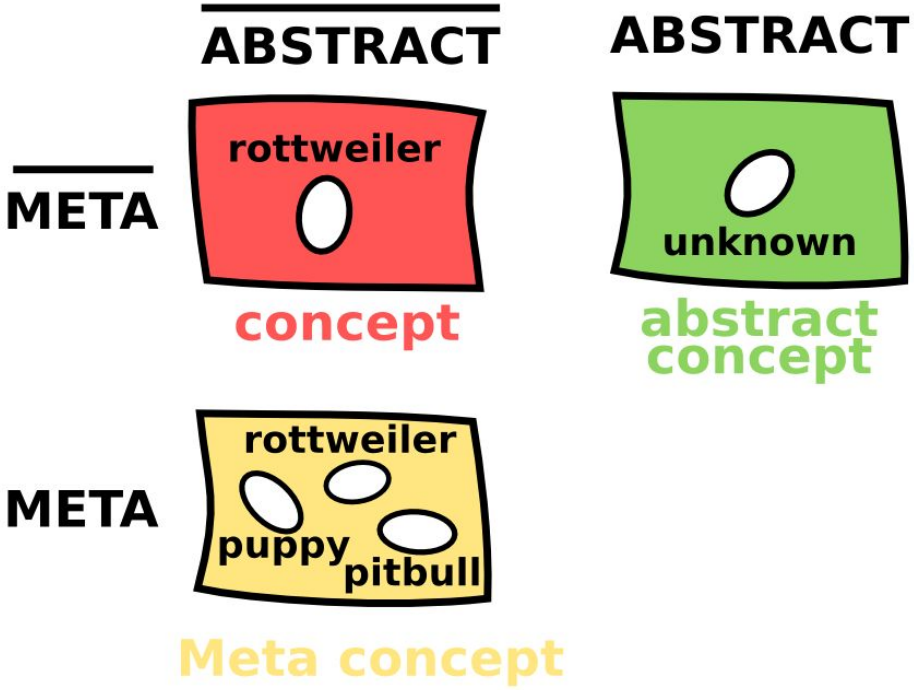
AMECON principle

- **AMECON: Abstract Meta-CONcept**
  - Abstract-concept + Meta-concept



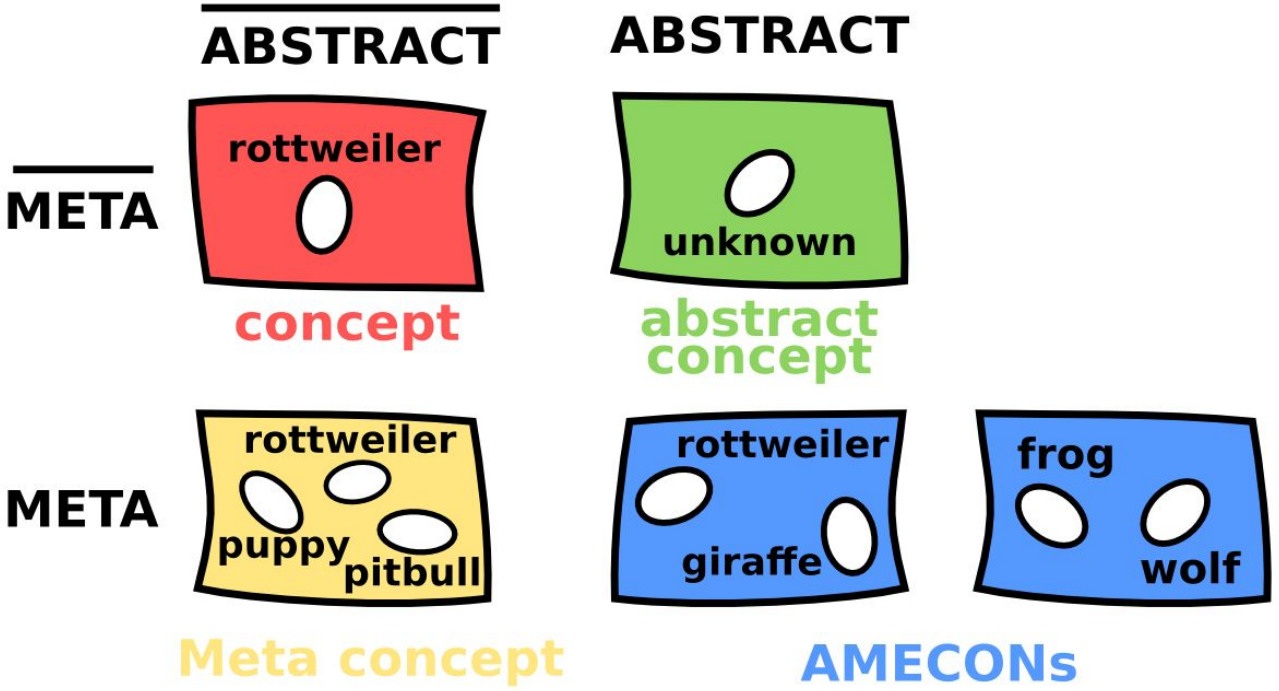
AMECON principle

- **AMECON: Abstract Meta-CONcept**
  - Abstract-concept + Meta-concept

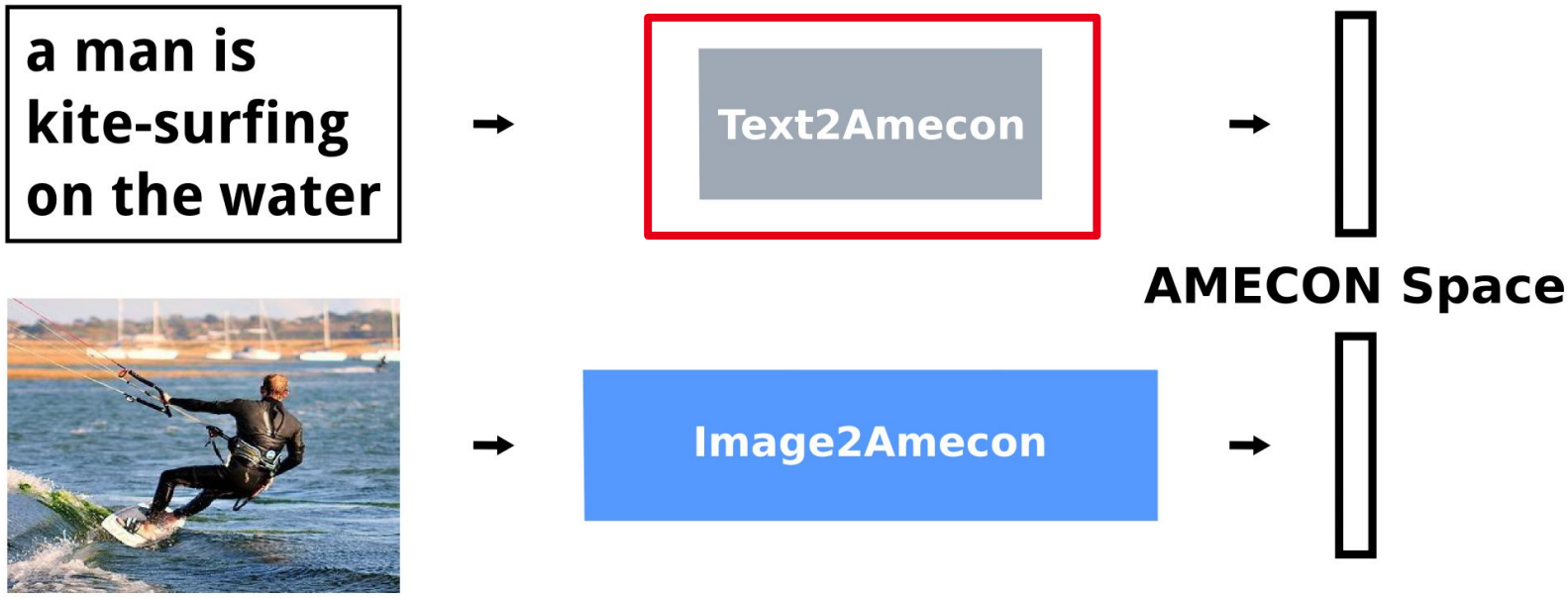


AMECON principle

- **AMECON: Abstract Meta-CONcept**
  - Abstract-concept + Meta-concept

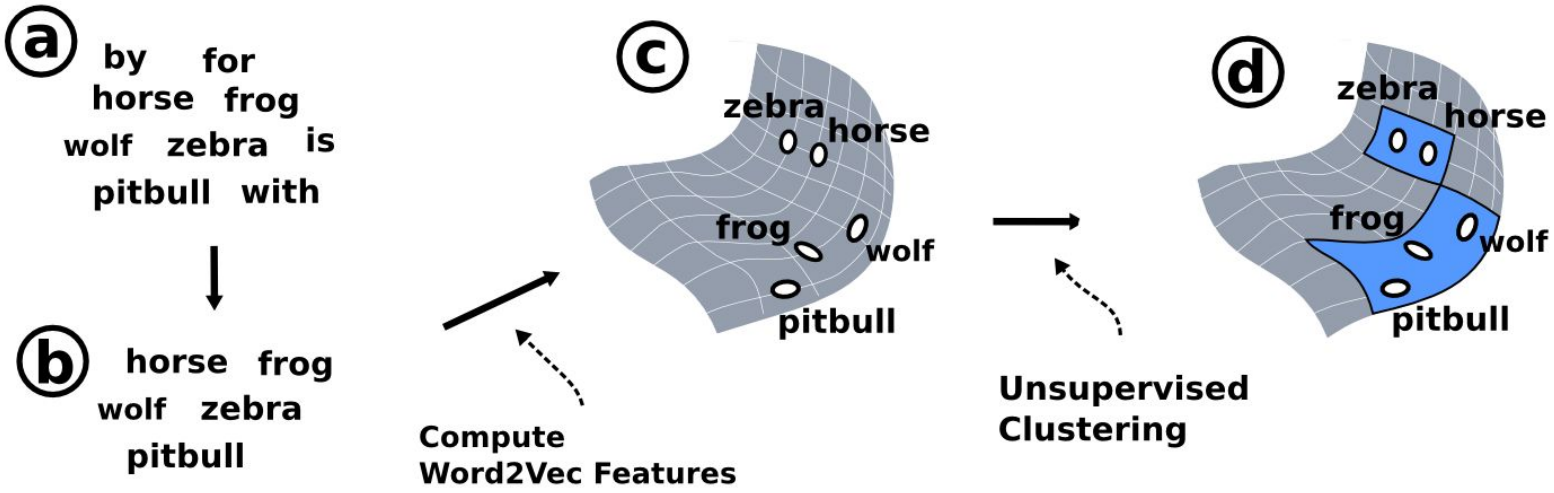


Overview of Our Approach



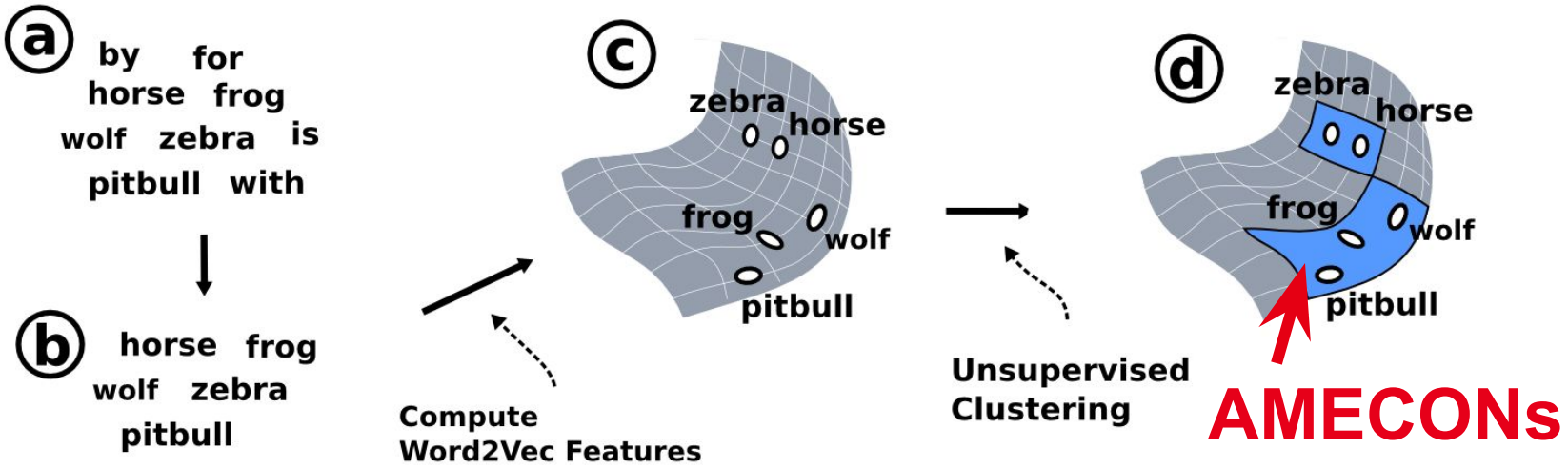
Learning Text2Amecon block

- **Learning Textual Features**
  a. Select all different words from training-data
  b. Remove stop-words (``is'', ``of'', ``for'', etc.)
  c. Compute word2vec features for each word
  d. Cluster (k-means) the whole set of features



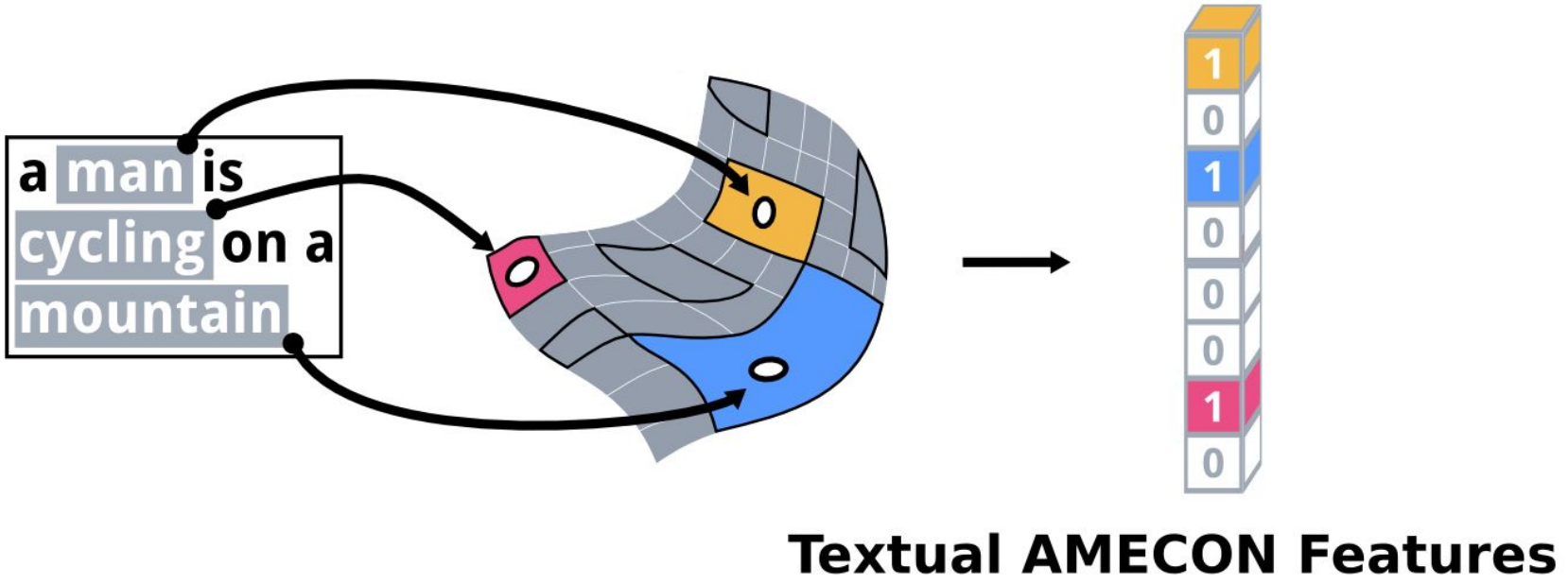
Learning Text2Amecon block

- **Learning Textual Features**
  a. Select all different words from training-data
  b. Remove stop-words (``is'', ``of'', ``for'', etc.)
  c. Compute word2vec features for each word
  d. Cluster (k-means) the whole set of features

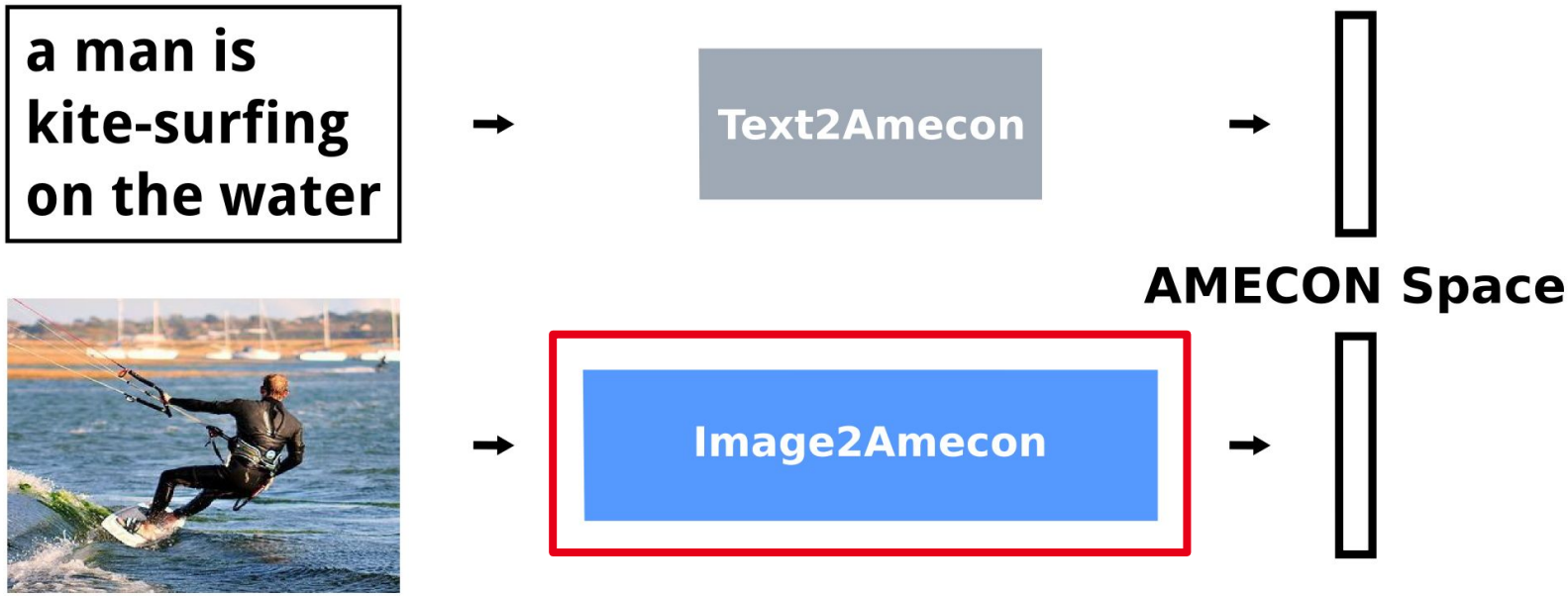


Computing Textual AMECON Features

Test phase



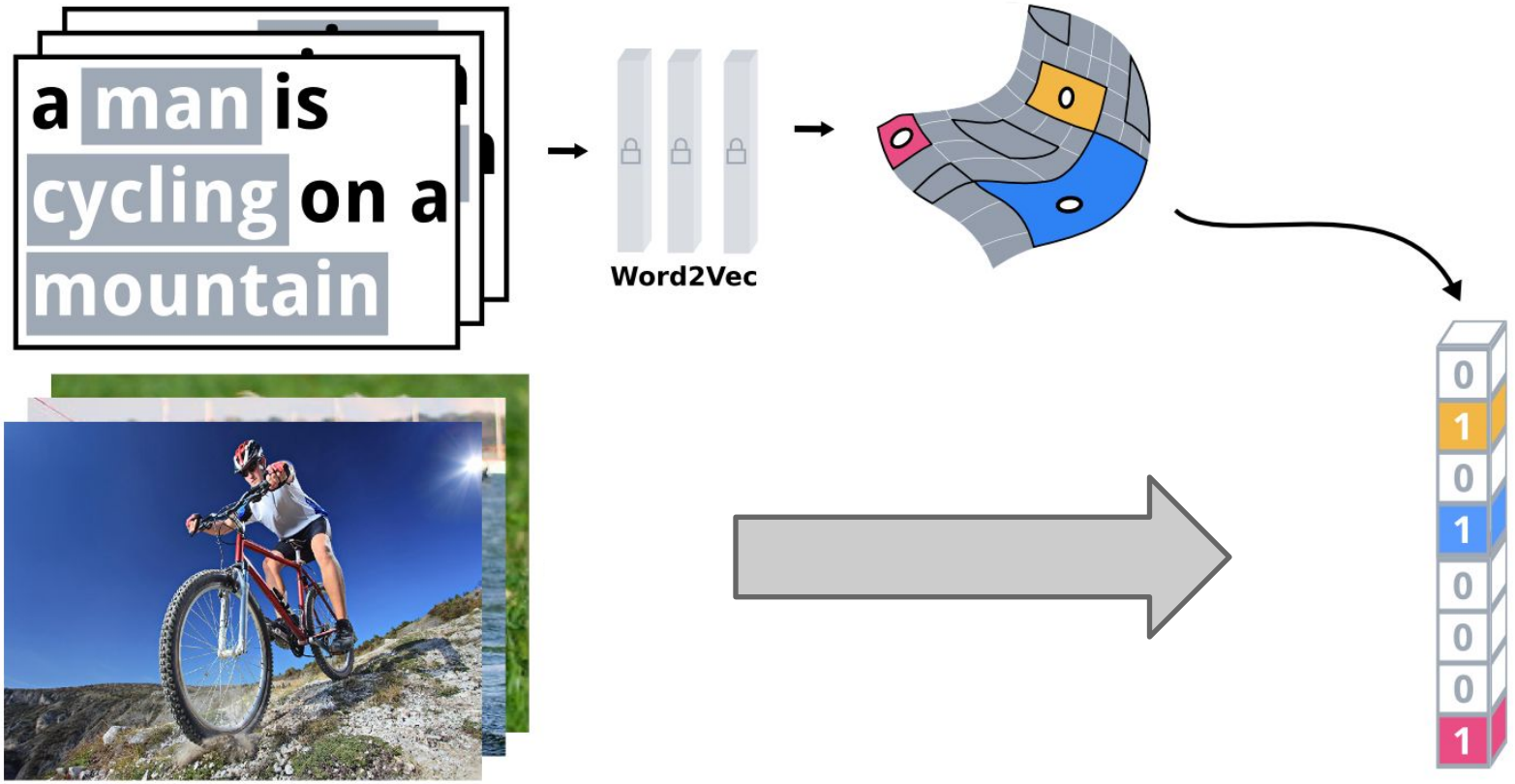
Overview of Our Approach



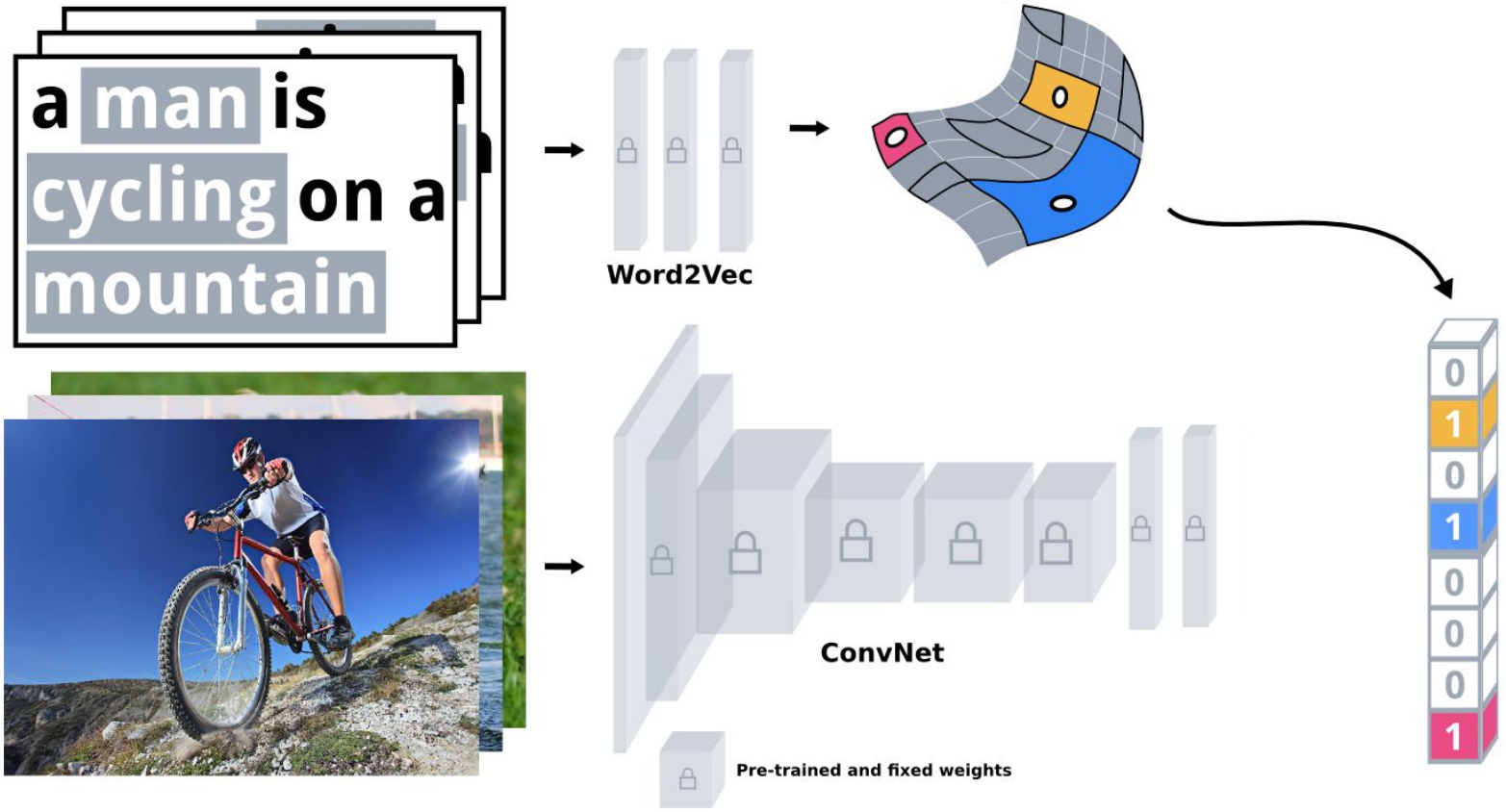
Learning Image2Amecon block

Learning Image2Amecon block

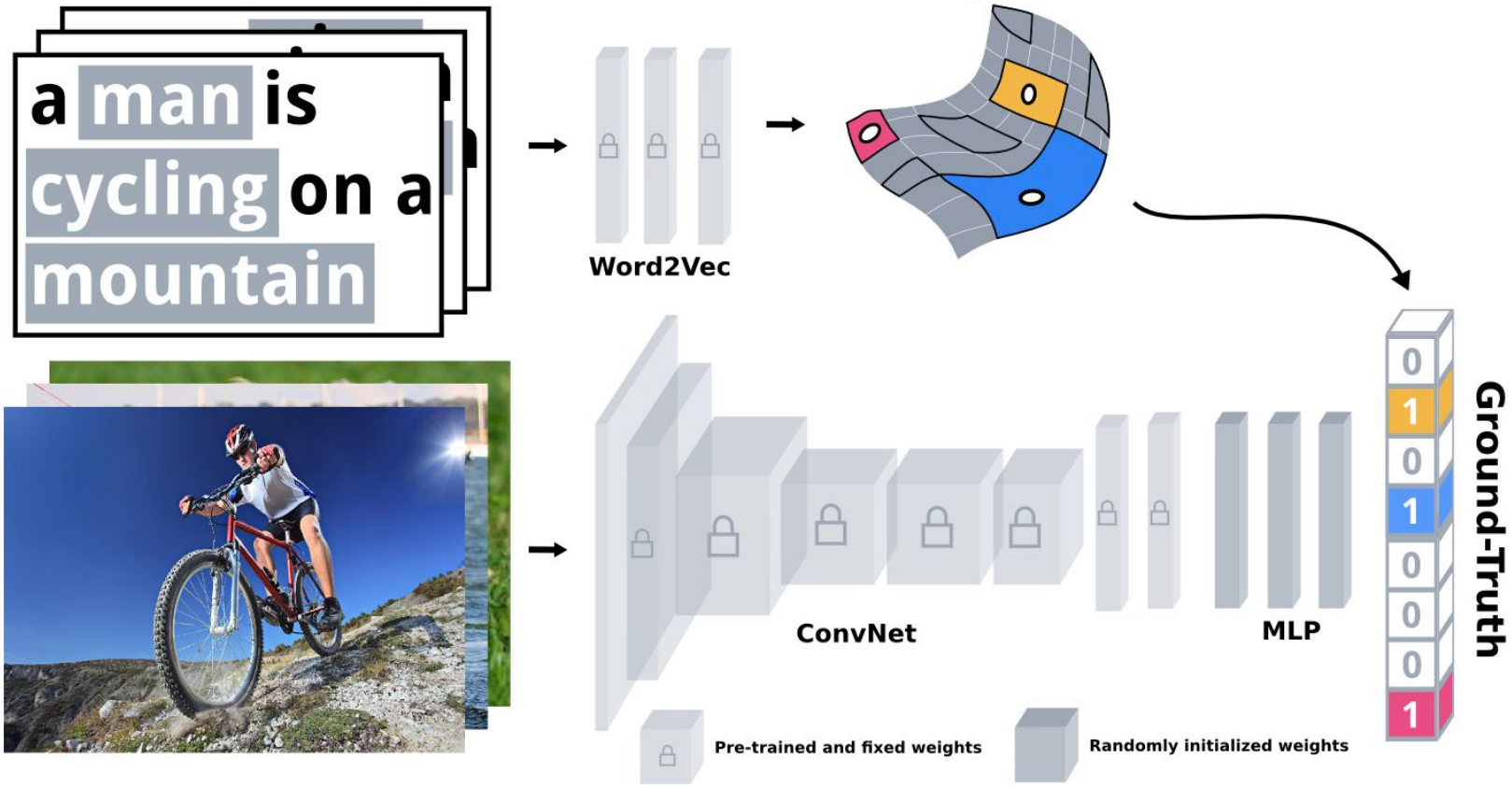
Learning Image2Amecon block



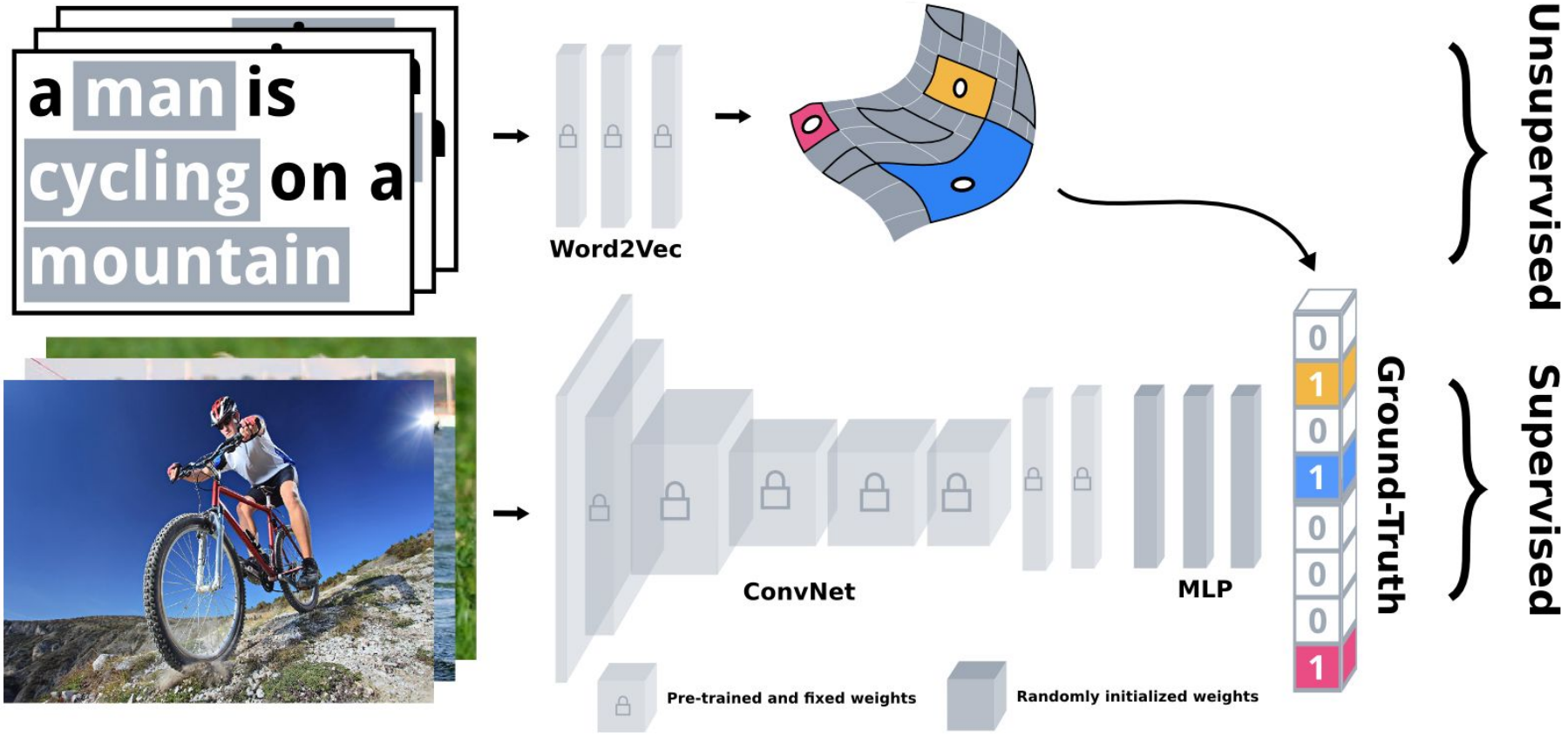
Learning Image2Amecon block



Learning Image2Amecon block

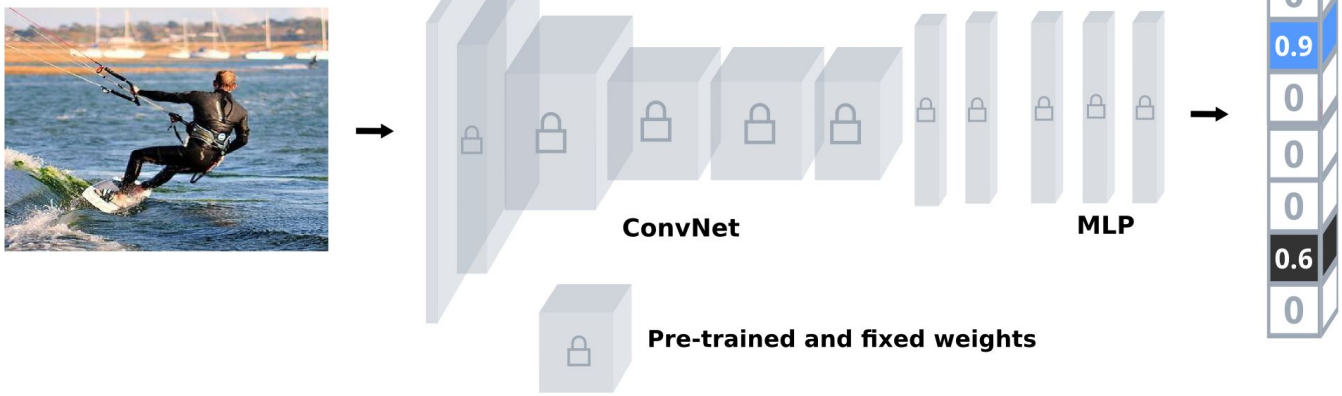


Learning Image2Amecon block



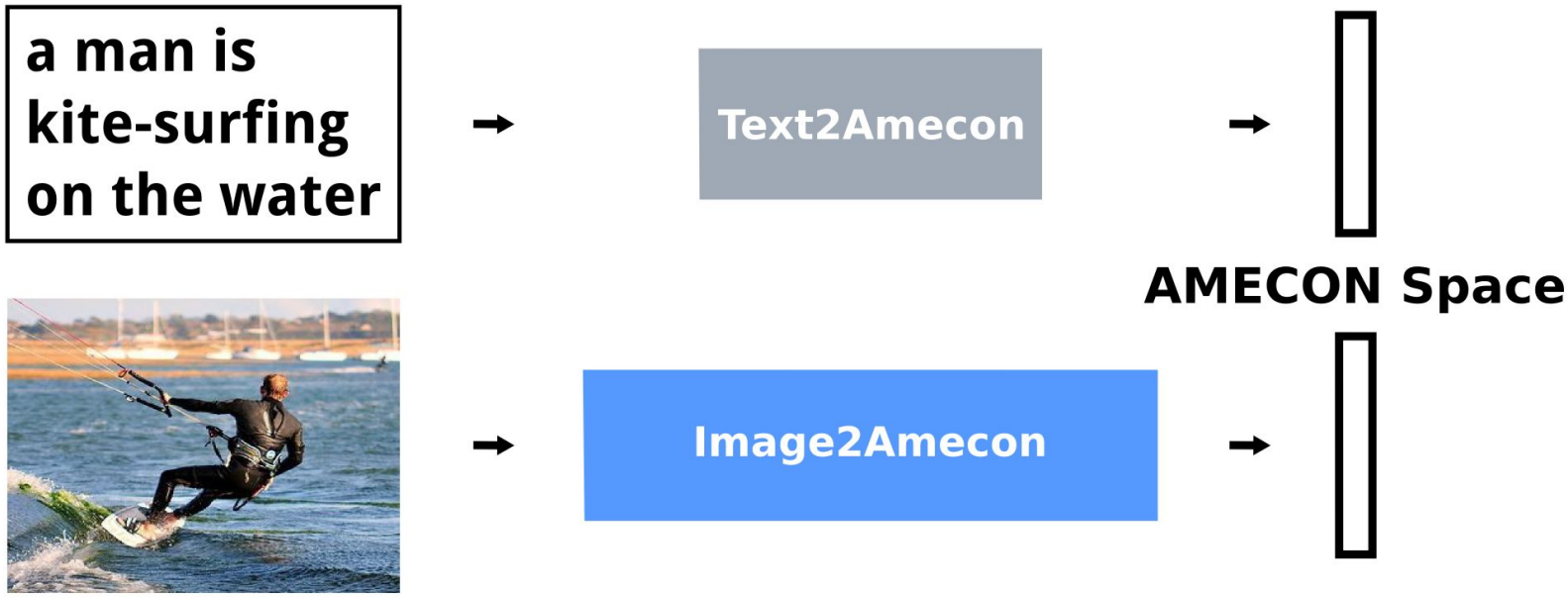
Computing Visual AMECON Features

Test phase



Visual AMECON Features

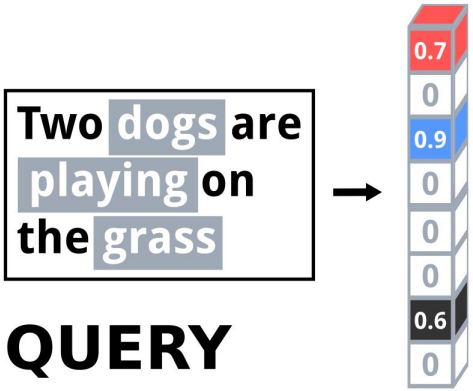
Overview of Our Approach



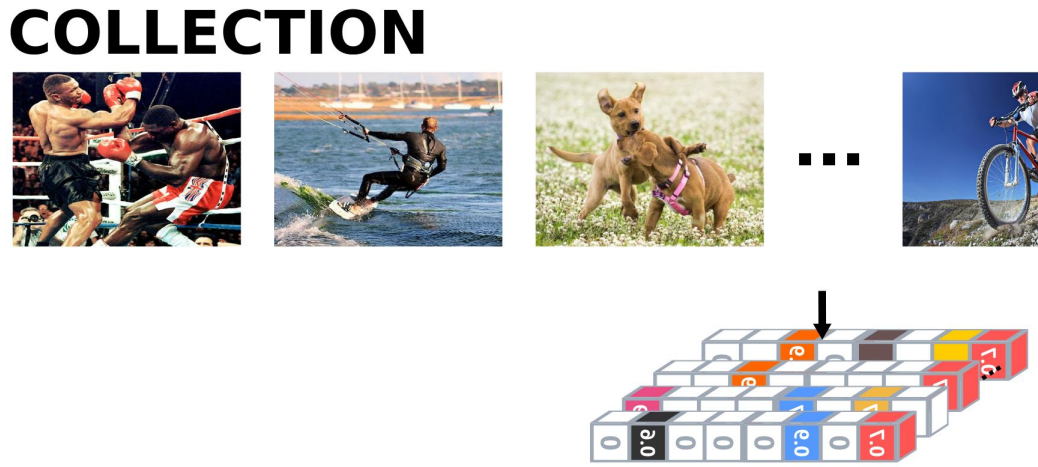
Matching Multi-Modal Data in AMECON Space

- **Matching texts & images in the same AMECON Space**
  - Text and Images directly comparable
  - Perform ANY multi-modal task

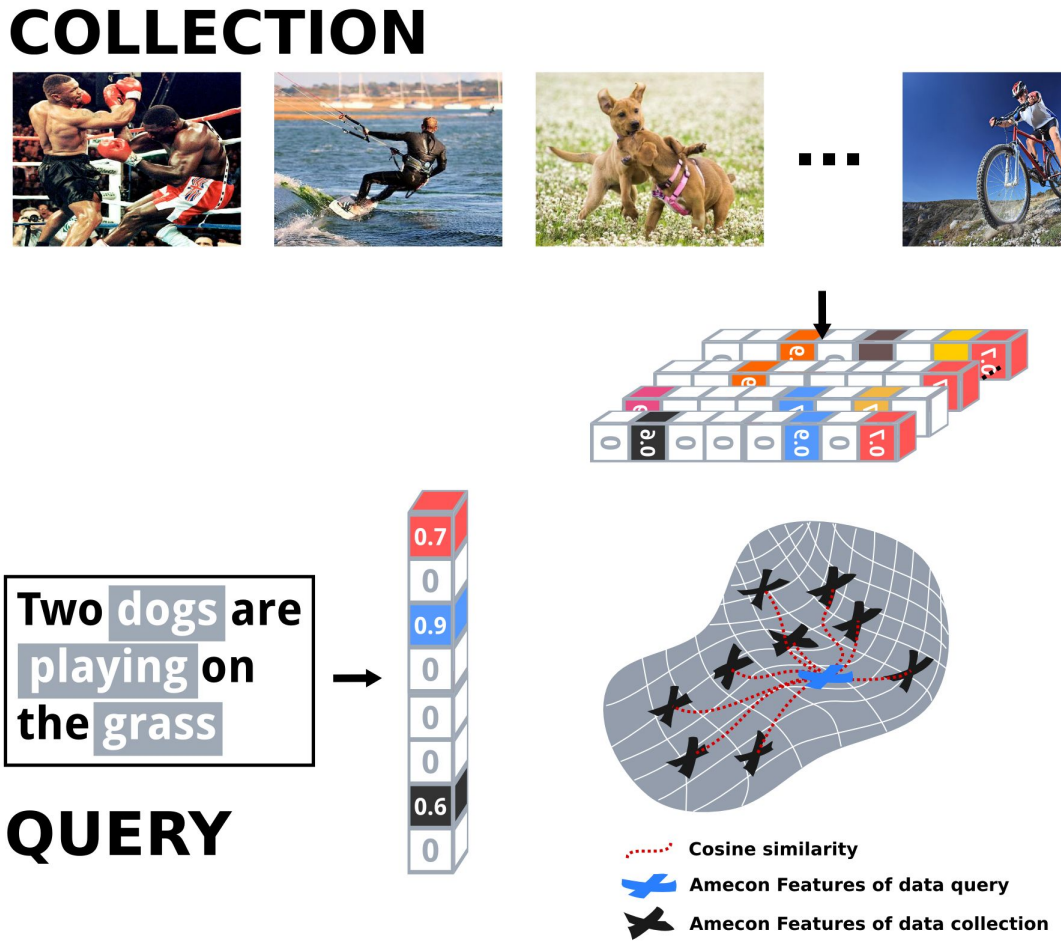
Text-Illustration in AMECON Space



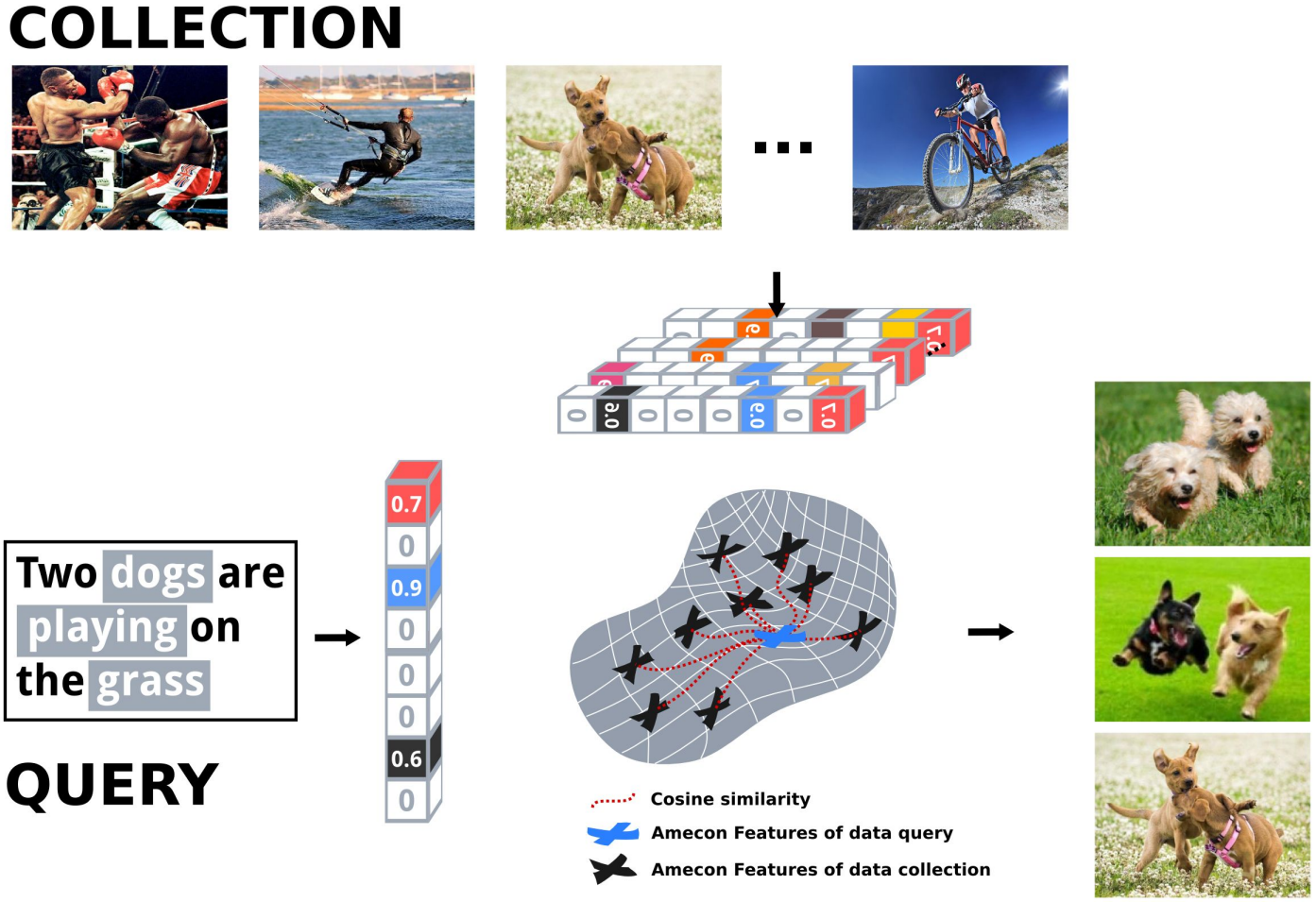
Text-Illustration in AMECON Space



Text-Illustration in AMECON Space



Text-Illustration in AMECON Space



Experimental Protocol

- **Training data**
  - 6,000/30,000 images in Flickr-8k/Flickr-30k
  - Each image associated to 5 captions

- **Testing data** (same for Flickr-8k & 30k)
  - 1000 images and 5000 captions
  - All captions as data-queries
  - All images as data-collection
  - Evaluation metric: Recall@K (K = 1, 5, 10)

Text Illustration Results

Method	Denotation	Flickr-8k			Flickr-30k		
		R@1	R@5	R@10	R@1	R@5	R@10
Karpathy *et al.* [17]	DeFrag						
Kiros *et al.* [18]	MNLM						
Mao *et al.* [21]	m-RNN						
Karpathy *et al.* [16]	BRNN*						
Yan *et al.* [36]	DCCA						
Tran *et al.* [32]	MACC†						
Our Approach	AMECON						

Neural Network-based Approach

CCA-based Approach

Our Approach

Text Illustration Results

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			10.3		
Kiros *et al.* [18]	MNLM	10.4			11.8		
Mao *et al.* [21]	m-RNN	11.5			12.6		
Karpathy *et al.* [16]	BRNN*	11.8			15.2		
Yan *et al.* [36]	DCCA	12.7			12.6		
Tran *et al.* [32]	MACC†	10.2			12.1		
Our Approach	AMECON	**15.9**			**18.3**		

Neural Network-based Approach

CCA-based Approach

Our Approach

Text Illustration Results

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		10.3	31.4	
Kiros *et al.* [18]	MNLM	10.4	31.0		11.8	34.0	
Mao *et al.* [21]	m-RNN	11.5	31.0		12.6	31.2	
Karpathy *et al.* [16]	BRNN*	11.8	32.1		15.2	37.7	
Yan *et al.* [36]	DCCA	12.7	31.2		12.6	31.0	
Tran *et al.* [32]	MACC†	10.2	29.3		12.1	33.5	
Our Approach	AMECON	**15.9**	**37.9**		**18.3**	**41.3**	

Neural Network-based Approach

CCA-based Approach

Our Approach

Text Illustration Results

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
Tran *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**

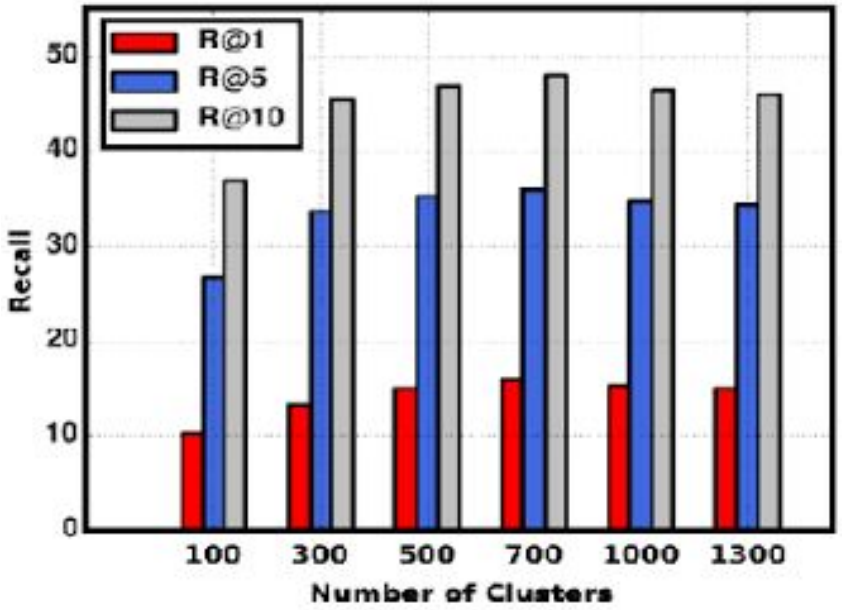
Neural Network-based Approach

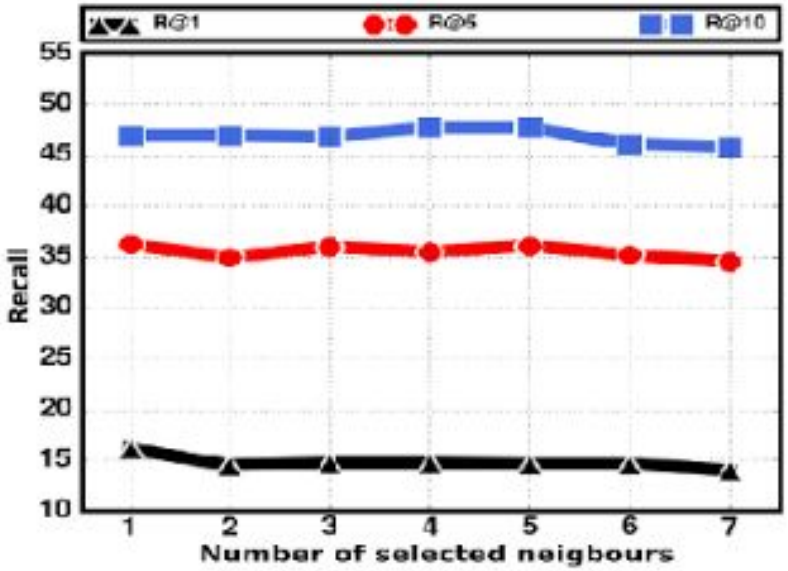
CCA-based Approach

Our Approach

Analysis of Parameters

- Quite robust to the parameters
  - Robust to #selected neighbours
  - Sensitive to #clusters (C) but stable when for a large range of C values





Conclusion

- **Novelty:**
  - Principle of AMECONs
    - Abstract MEta-CONcepts
  - Mixing supervised and unsupervised learning to build a multi-modal space
- **Results on Text-illustration:**
  - +4 points of R@K (avg.) compared to best methods of the literature
- **Future Work:**
  - Image captioning with AMECON-features

Code will be released at:

http://perso.ecp.fr/~tamaazouy/

Thank you (questions ?)

Commissariat à l'énergie atomique et aux énergies alternatives
Institut List | CEA SACLAY NANO-INNOV | BAT. 861 – PC142
91191 Gif-sur-Yvette Cedex - FRANCE
www-list.cea.fr