What is Happening: Annotating Images with Verbs

Author: Gang Tian, gtia8901@uni.sydney.edu.au; Genliang Guan, genliang.guan@sydney.edu.au
Supervisor: Dr Zhiyong Wang & Prof. David Feng
School of Information Technologies
FACULTY OF ENGINEERING & INFORMATION TECHNOLOGIES

1. INTRODUCTION

Background

- "A picture is worth 1000 words", image semantic understanding and description are still important challenges.
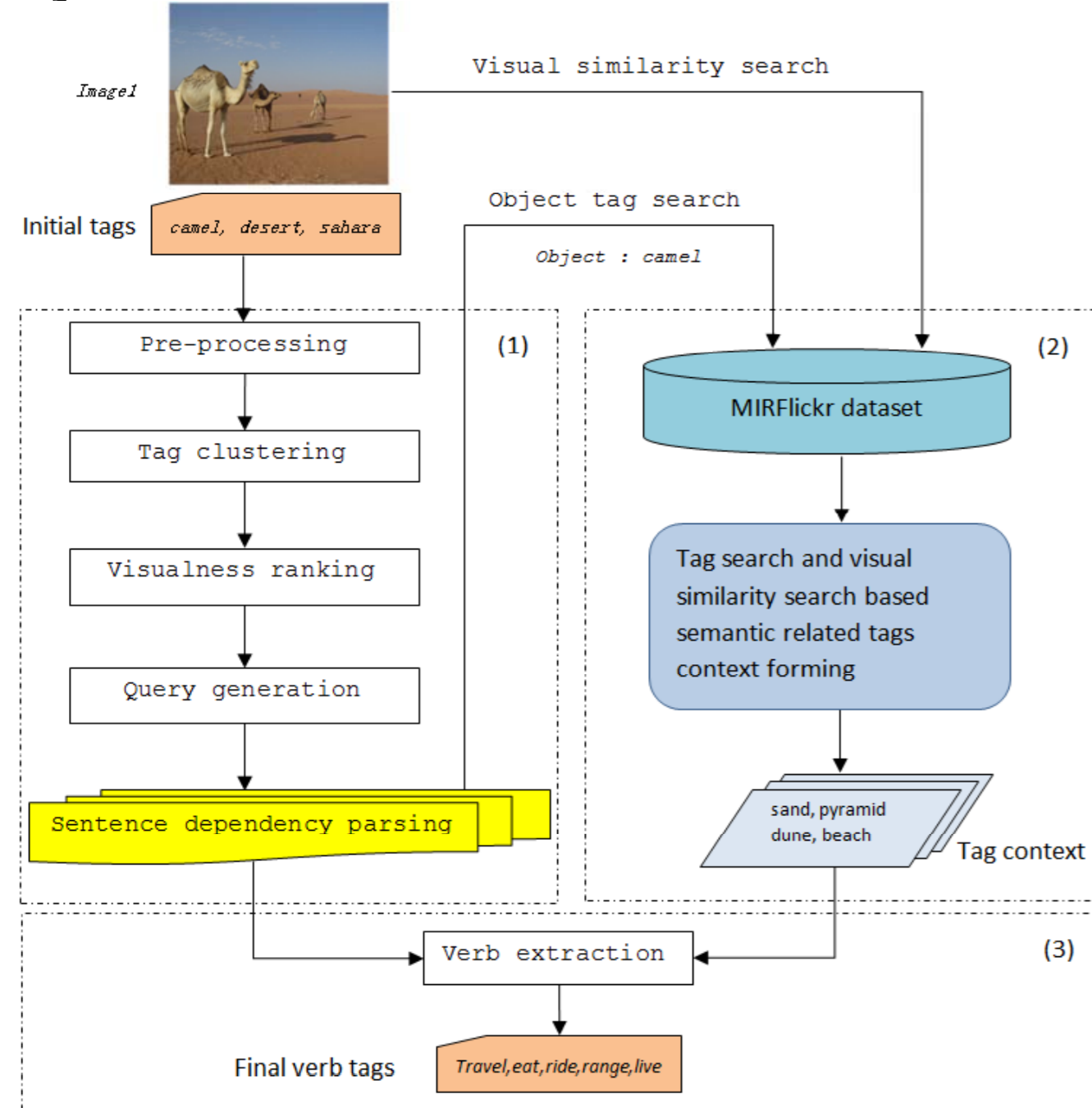
Problem

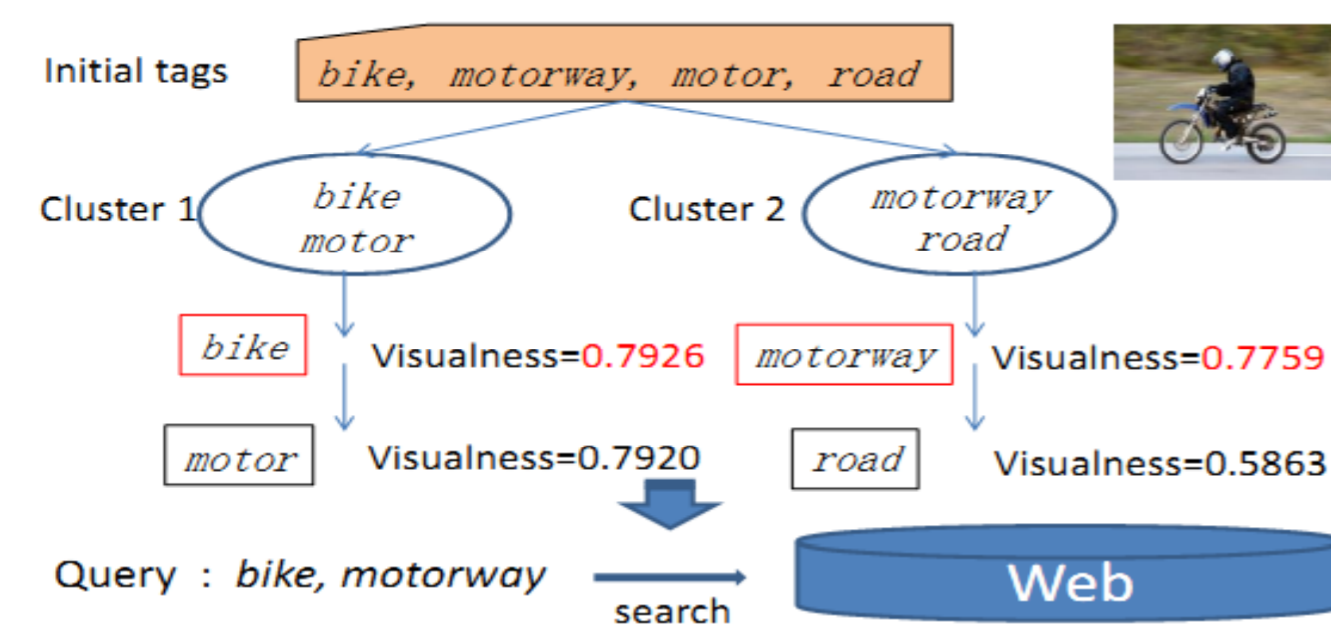- Most image annotation works focus on noun tags (concepts & objects)

Aim

- Generating verb tags which can enable richer image semantic understanding.
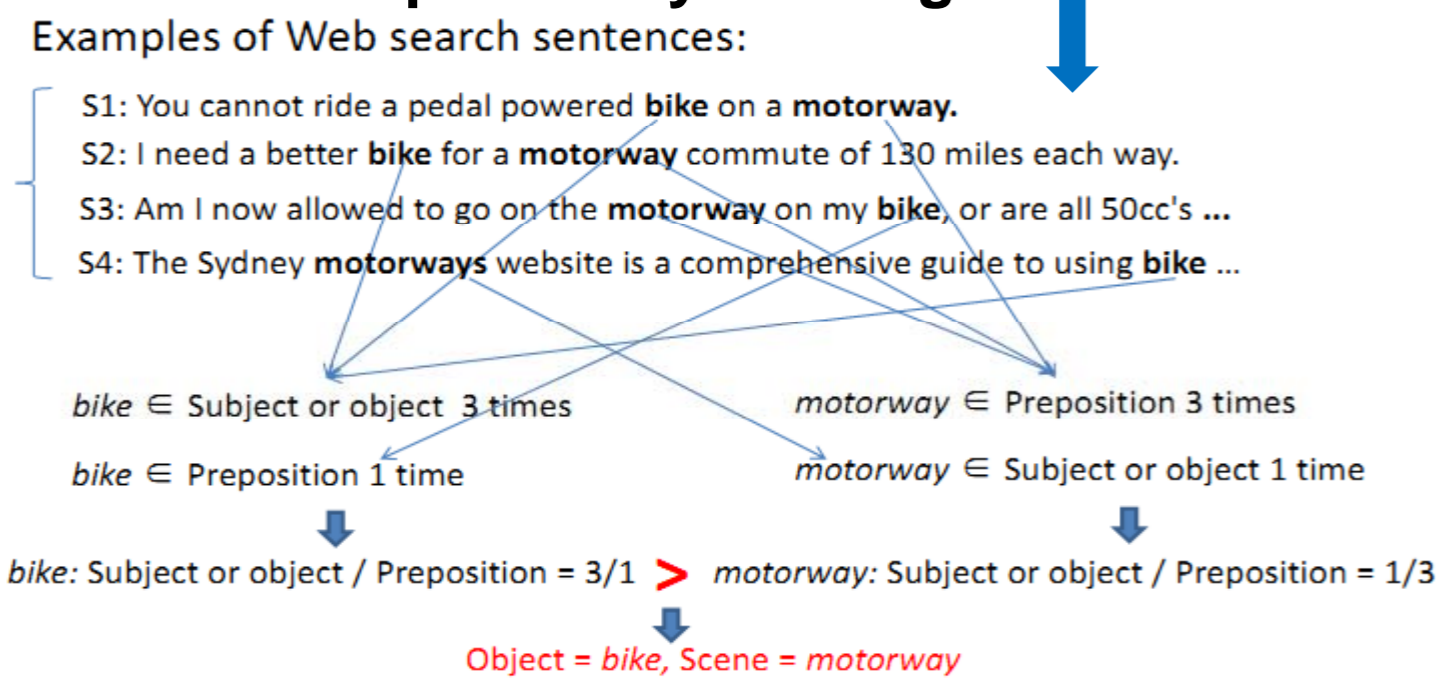
2. METHODOLOGY

System Framework

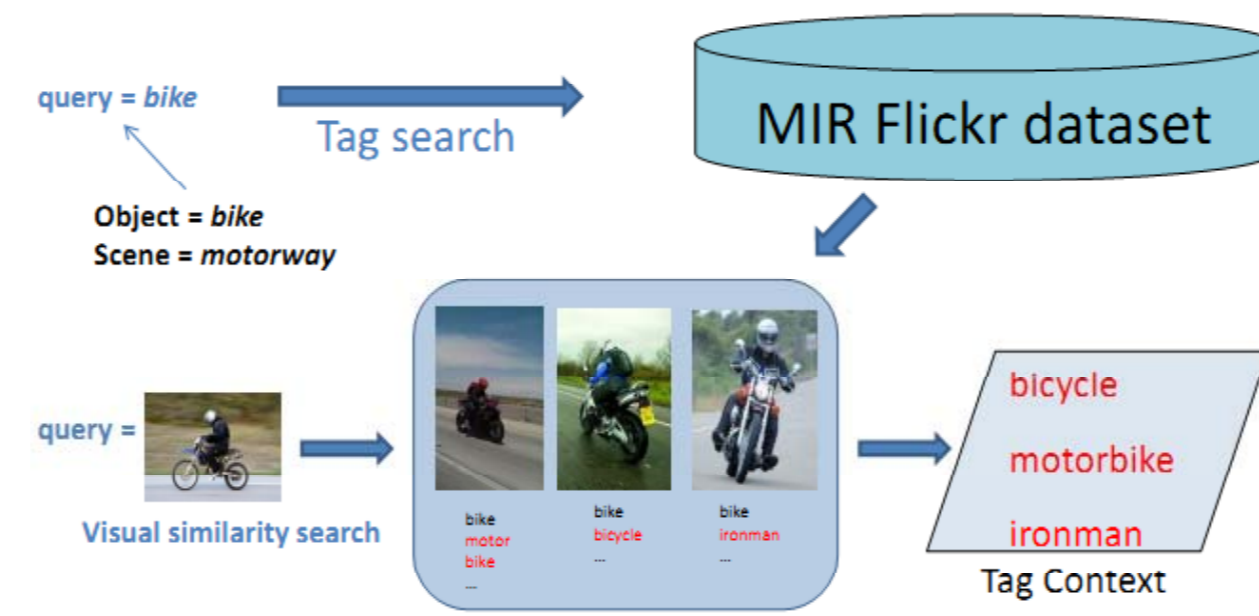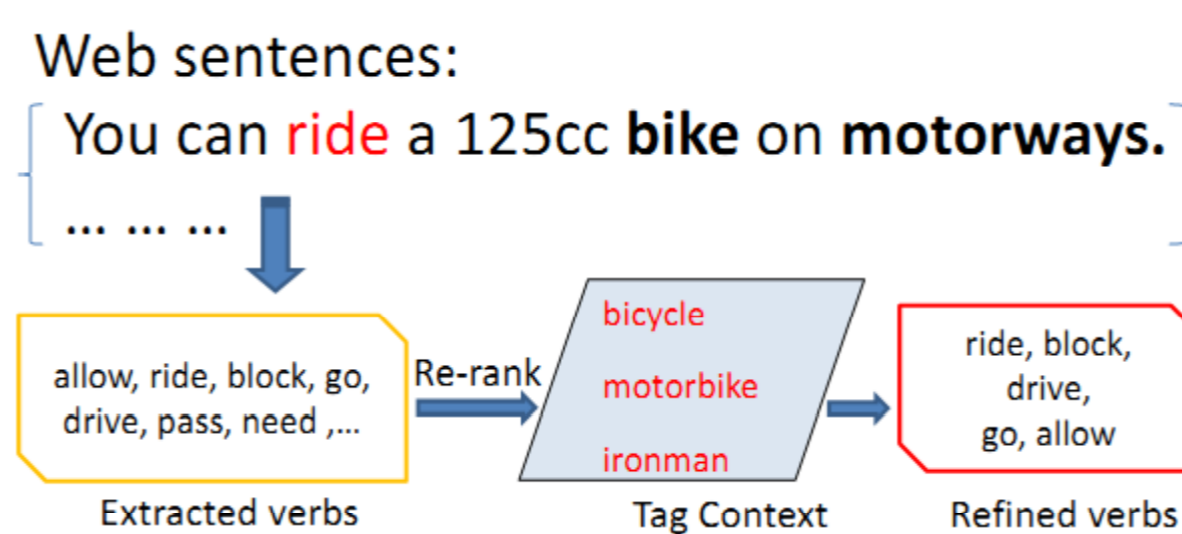### Tag Clustering & Visualness Ranking

Sentence Dependency Parsing

Examples of Web search sentences:

S1: You cannot ride a pedal powered **bike** on a **motorway.**
S2: I need a better **bike** for a **motorway** commute of 130 miles each way.
S3: Am I now allowed to go on the **motorway** on my **bike**, or are all 50cc's ...
S4: The Sydney **motorways** website is a comprehensive guide to using **bike** ...

bike ∈ Subject or object 3 times
bike ∈ Preposition 1 time
motorway ∈ Preposition 3 times
motorway ∈ Subject or object 1 time

bike: Subject or object / Preposition = 3/1 **>** motorway: Subject or object / Preposition = 1/3

Object = bike, Scene = motorway

Annotation Results

Image No.	Initial Tags	Context Tags	V2 Top5
	camel	sand	travel
	desert	pyramid	eat
	Sahara	dune	ride
		beach	range
Image1		travel	live

Image No.	Initial Tags	Context Tags	V2 Top5
	duck	water	show
	mallard	duckling	cripple
	bird	drake	wound
		male	rise
Image2		flight	lure

Image No.	Initial Tags	Context Tags	V2 Top5
	lion	cat	kill
	zoo	baby	issue
		panther	deny
		dandelion	tear
Image3		cub	live

Image No.	Initial Tags	Context Tags	V2 Top5
	banjo	musician	play
		guitar	blend
		hat	try
		instrument	associate
Image4		fiddle	

Image No.	Initial Tags	Context Tags	V2 Top5
	bike	bicycle	ride
	motor	motorbike	block
	motorway	ironman	drive
	road	race	go
Image5		cycling	allow

Image No.	Initial Tags	Context Tags	V2 Top5
	bus	transport	think
	street	car	run
		travel	block
		transit	show
Image6		trolleybus	wait

Image No.	Initial Tags	Context Tags	V2 Top5
	child	woman	slippered
	girl	female	witness
	slipper	lady	Slipper
		teen	happen
Image7		beach	carry

Image No.	Initial Tags	Context Tags	V2 Top5
	bird	duck	glide
	lake	seagull	move
		flight	breed
		beach	visit
Image8		feather	populate

Image No.	Initial Tags	Context Tags	V2 Top5
	airbus	airplane	land
	plane	airliner	crash
	airport	aircraft	help
		jet	refuel
Image9		aeroplane	fly

Image No.	Initial Tags	Context Tags	V2 Top5
	tree	snow	grow
	bank	branch	root
	sky	sun	remember
		park	connect
Image10		countryside	caution

Extracting and Re-ranking Verbs

Extracting and Re-ranking Verbs

Web sentences:

You can **ride** a 125cc **bike** on **motorways.**
... ...

allow, ride, block, go, drive, pass, need ,... → Re-rank → bicycle / motorbike / ironman → ride, block, drive, go, allow

Extracted verbs | Tag Context | Refined verbs

3. EXPERIMENTS

Experiment Setting

- Dataset: MIR Flickr (http://press.liacs.nl/mirflickr/)
- Ground-truth data: manually annotated by 11 assessors

race	11
drive	8
ride	8
go	6
pass	4
wear	4
run	3
travel	3
compete	2
crash	2

ride — 5: strongly relevant
block — 1
drive — 5
go — 4
allow — 1: irrelevant

Generated verbs (ranking by frequency) | Score

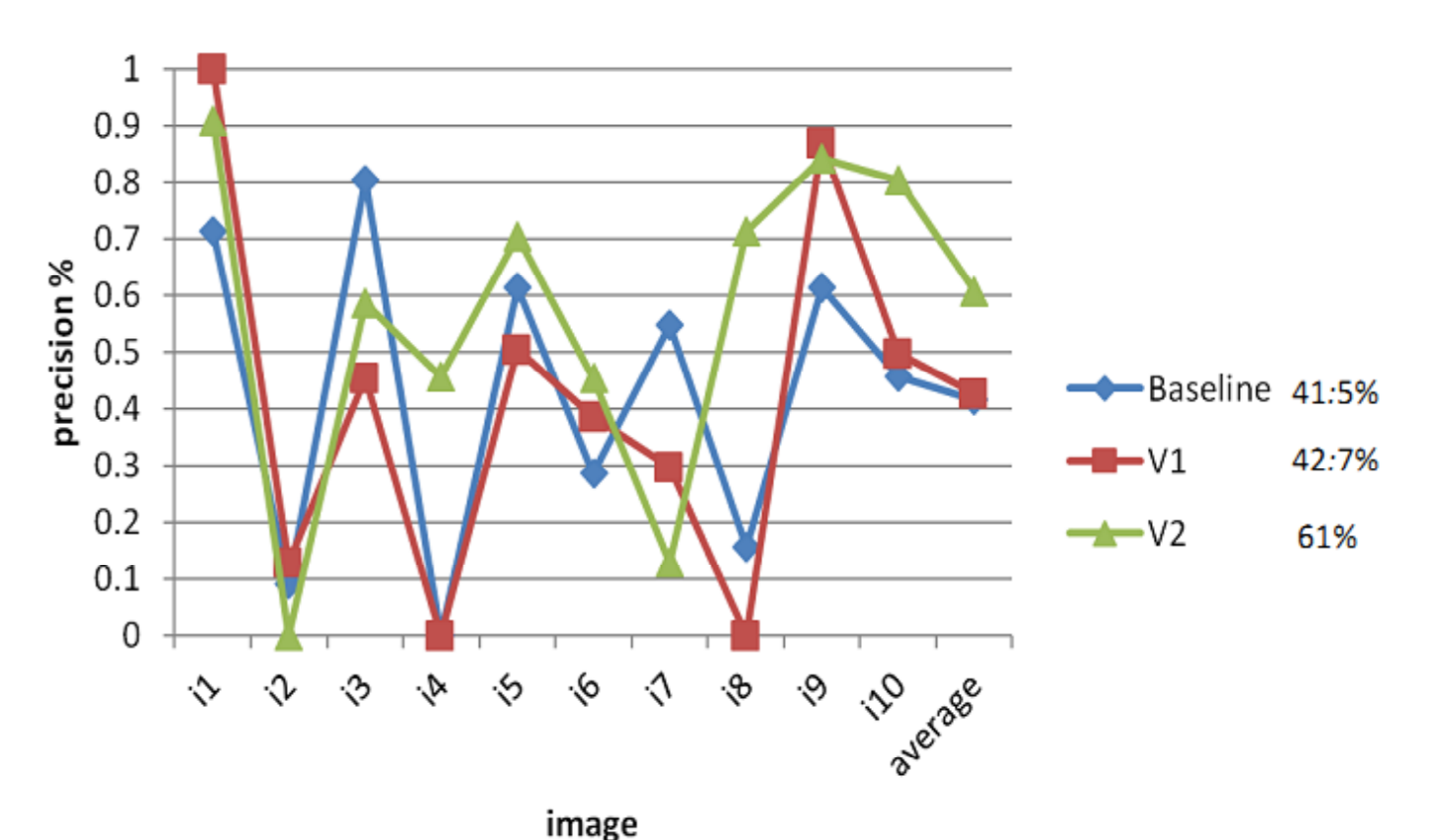Top 10 annotated verb terms and frequency

Comparison Methods

- Baseline : each initial tag as query for web search
- Method V1 : object and scene tags based web search
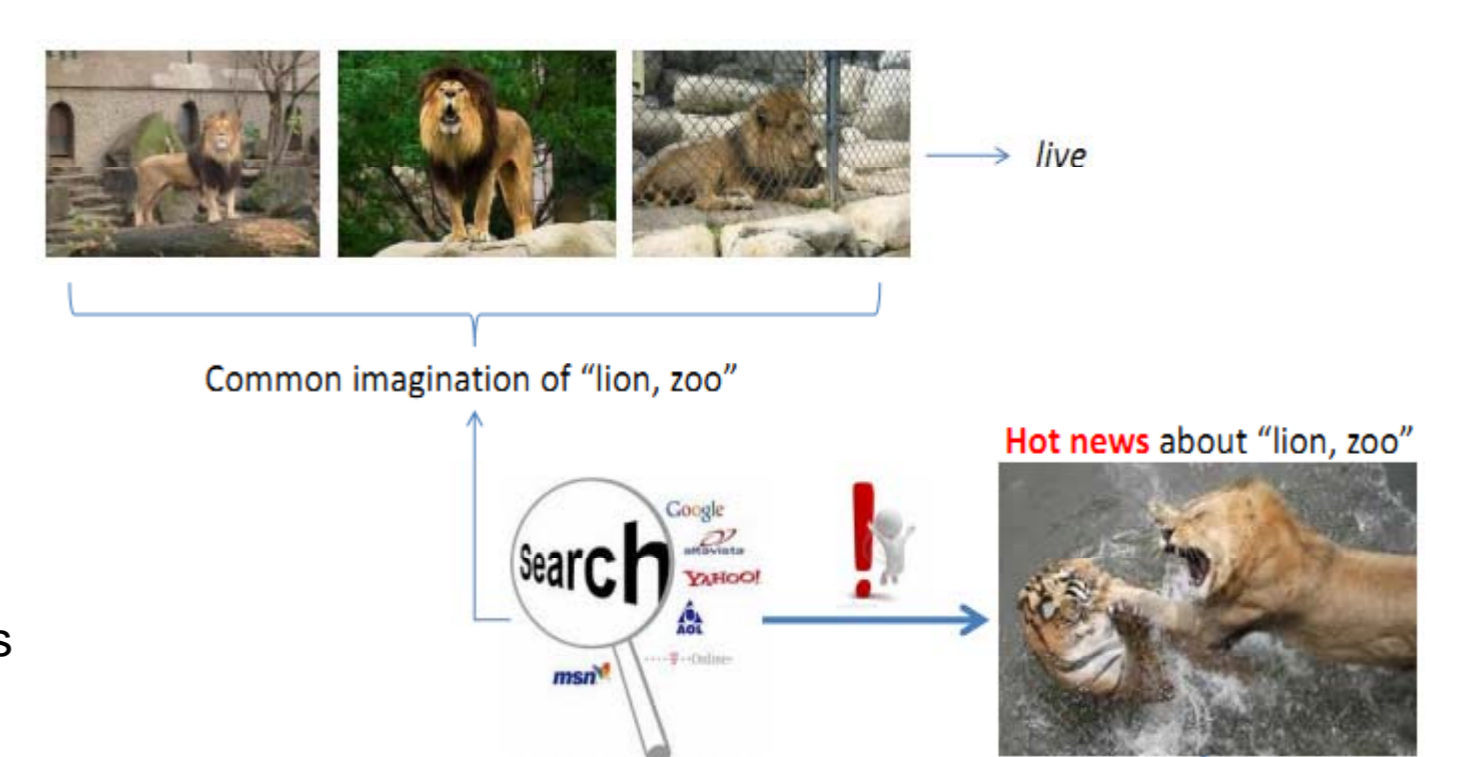- Method V2 : sentence dependency structure & tag context based re-rank

Comparison (Relevance)

V2 > V1 > Baseline

Tags clustering (Object/Scene) is useful.

Sentence dependency parsing & tag context based re-rank is critical.
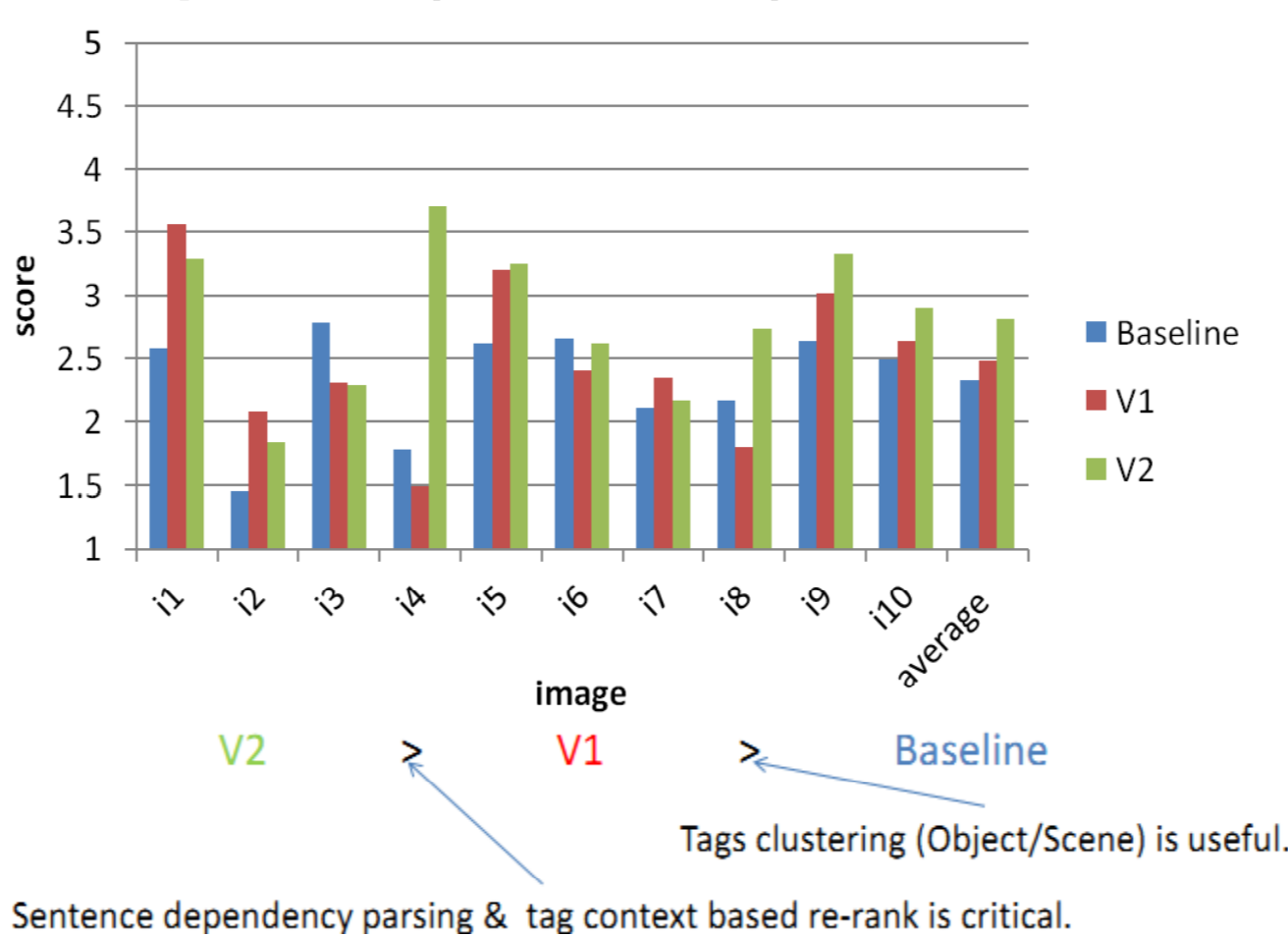
Comparison against the Ground-truth

Baseline 41:5%
V1 42:7%
V2 61%

Impact of Web Search

Common imagination of "lion, zoo" → live

Hot news about "lion, zoo"
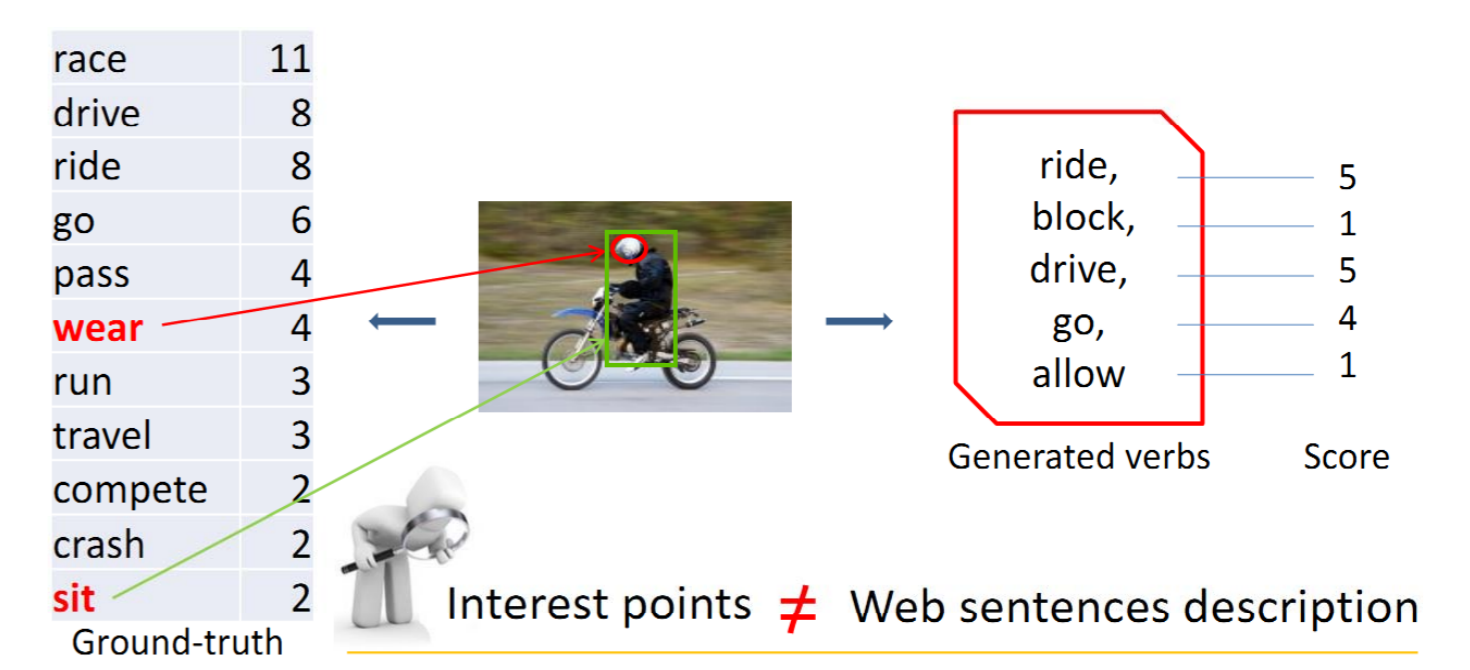
"The zoo in the Turkish capital of Ankara issued a statement Monday to deny media reports that the tiger **bent** iron bars to **attack** the lion and **tear** it into pieces in the attack ..."

Diversity of Human Perception

race	11
drive	8
ride	8
go	6
pass	4
wear	4
run	3
travel	3
compete	2
crash	2
sit	2

Ground-truth

ride, block, drive, go, allow — 5, 1, 5, 4, 1

Generated verbs | Score

Interest points ≠ Web sentences description

4. CONCLUSIONS

- Verb tags discovering: the underlying activities for given still image
- Context: potential textual information from the Web sentence corpus & visual corpus by CBIR
- Future work: advanced context modeling algorithms to refine verb candidates

5. REFERENCES

- Y. Ushiku, T. Harada, and Y. Kuniyoshi. Automatic sentence generation from images. ACM Multimedia, 2011.

This work has been published in the 20th Anniversary ACM Multimedia 2012, Nara, Japan.