

# DESCRIBING VERBS USING VISUAL VECTORS

IRENE SUCAMELI

PHD IN



AT



# OUTLINE

1. Verbs + images, what are we talking about?
2. A gentle introduction to Distributional Semantics Models
3. Our research
4. Preliminary results
5. Conclusion

# PART 1: WHAT ARE WE TALKING ABOUT?

YOU SHALL KNOW A  
WORD BY THE  
COMPANY IT KEEPS


(FIRTH 1957: 11)

YOU SHALL KNOW A  
WORD BY THE  
COMPANY IT KEEPS

THE DISTRIBUTIONAL HYPOTHESIS:

LEXEMES WITH SIMILAR DISTRIBUTIONAL PROPERTIES HAVE SIMILAR MEANINGS.

YOU SHALL KNOW A  
VERB BY THE  
IMAGES IT KEEPS

A photograph of two Super Mario Bros. figurines, Luigi and Mario, standing on a dirt path. Luigi is on the left, wearing his signature green hat with a white 'L' and blue overalls. Mario is on the right, wearing his red hat with a white 'M' and blue overalls. Both have orange noses and black mustaches. A white speech bubble originates from Mario's head, containing the text 'Aww man, how can I save my beloved?'. The background is a blurred green forest. The bottom right corner of the image has a page number '5/30'.

Aww man, how can I  
**save** my beloved?





Aww man, how can I  
**save** my beloved?





# PART 2: DSMS

# A GENTLE INTRODUCTION TO DSMS

Distributional Semantic Models (DSMs) apply the distributional hypothesis to study semantic facts on a computational level.

...the best way to save **money** is contact a bank...

...in stories, the **princess** is saved by the hero...

...a lot of **money** were paid to rescue the hostages...

...who says that the **princess** cannot rescue the prince?...

# A GENTLE INTRODUCTION TO DSMS

Traditional DSMS use co-occurrence matrices to record how many times lexemes co-occur with selected contexts.

	<b>princess</b>	<b>money</b>	Lexical item == n-dimensional distributional vector whose components are distributional features representing its co-occurrences with linguistic contexts.
<b>save</b>	2	3	
<b>rescue</b>	3	2	

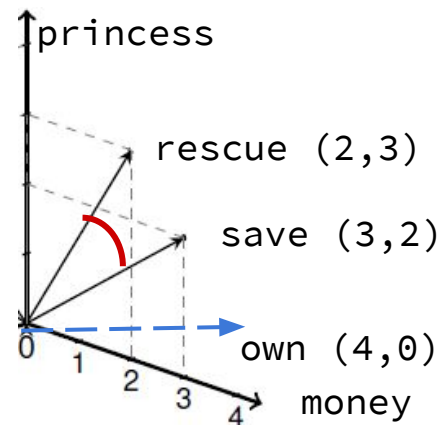
# A GENTLE INTRODUCTION TO DSMS

The distributional similarity between two lexemes  $u$  and  $v$  is measured with the similarity between their distributional vectors  $u$  and  $v$ .

Cosine similarity:

$$\cos(\vec{v}, \vec{w}) = \frac{\vec{v} \cdot \vec{w}}{|\vec{v}| |\vec{w}|} = \frac{\vec{v}}{|\vec{v}|} \cdot \frac{\vec{w}}{|\vec{w}|}$$

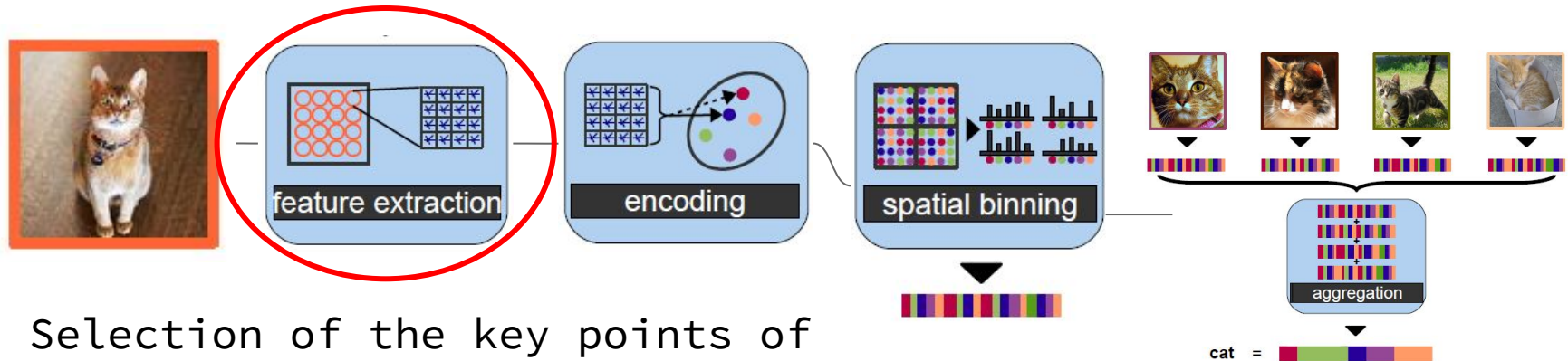
Range: from 1 to -1



**Cosine  
similarity:  
0.89**

# BoVW

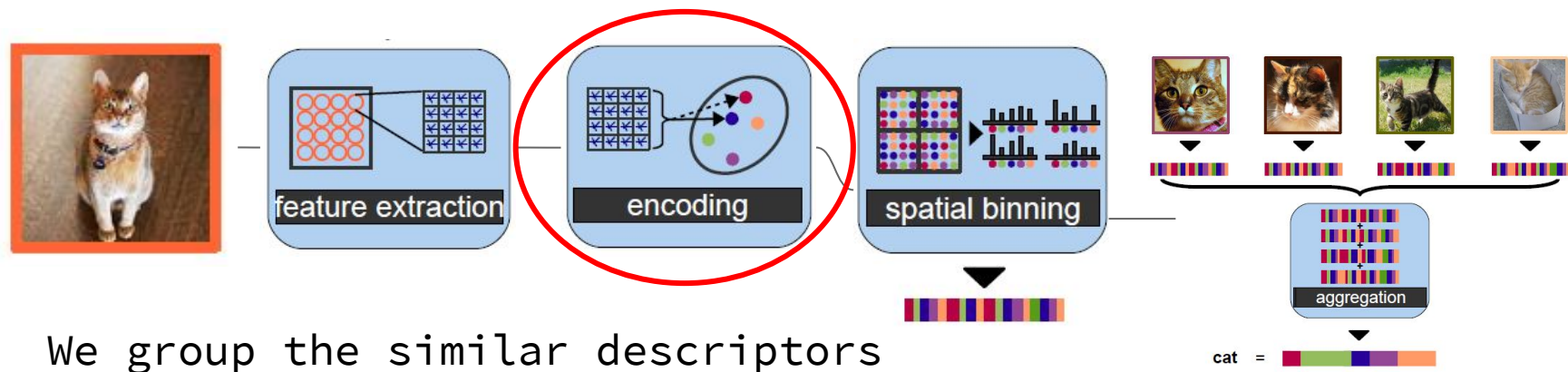
In visual distributional semantic models, words are described through their visual representations, aka visual words -> **Bag-of-Visual-Words**.



Selection of the key points of an image. We compute the SIFT *descriptors*

# BoVW

In visual distributional semantic models, words are described through their visual representations, aka visual words -> **Bag-of-Visual-Words**.

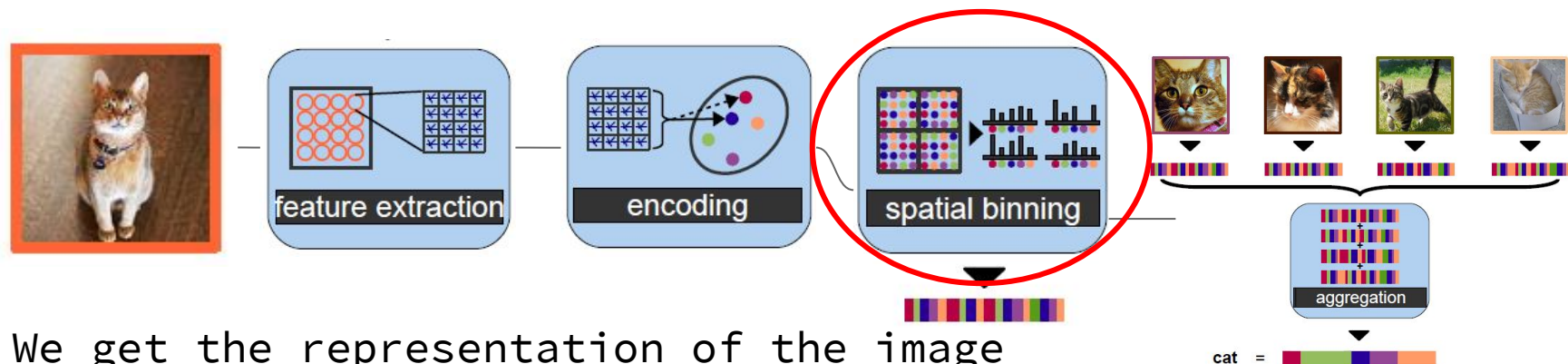


We group the similar descriptors



# BoVW

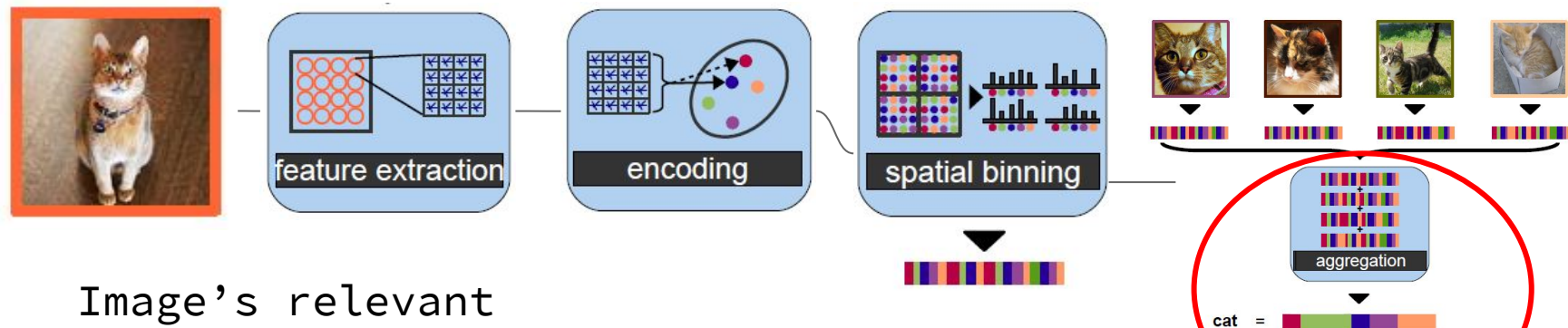
In visual distributional semantic models, words are described through their visual representations, aka visual words -> Bag-of-Visual-Words.



We get the representation of the image according to the vector distribution

# BoVW

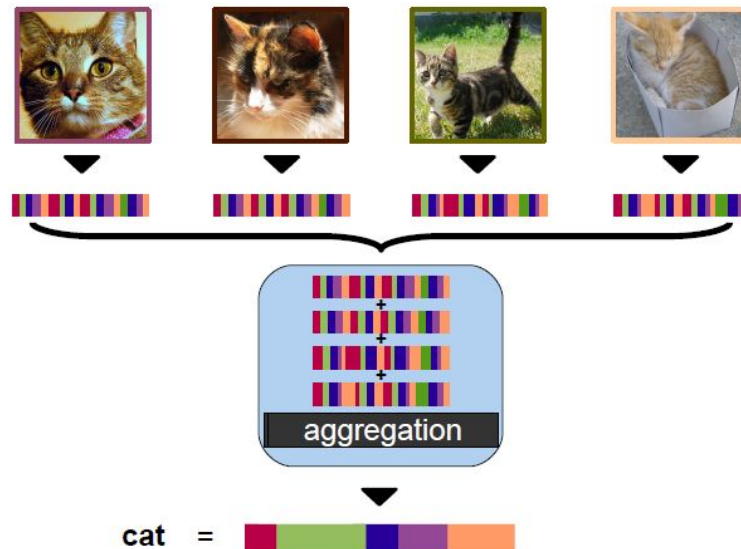
In visual distributional semantic models, words are described through their visual representations, aka visual words -> **Bag-of-Visual-Words**.



Image's relevant representations are combined

# BoVW

Given a set of images that describes the same concept, the representation of the concept is obtained from the sum of the vectors which describe the input images.



# PART 3 - OUR RESEARCH

# OUR RESEARCH

To demonstrate that we can describe the similarity between verbs using Distributional Visual Models we created and compared three models: two text-based ones and a visual one.

The vector of the verb is derived from the set of nouns that typically co-occur with that verb as subjects and objects.

$$\vec{V} = \vec{V}_{\text{Subj}} \oplus \vec{V}_{\text{Obj}}$$

# LINGUISTICS RESOURCES


100 verbs extracted from `SimLex-999` (Hill et al., 2015)

- 999 pairs of words for three POS;
- often used for the evaluation of DSMs;
- describes the similarity between pairs of words instead of their association score;
- antonyms are described as pairs with a low similarity score.



# LINGUISTICS RESOURCES

Nouns were extracted from the **Distributional Memory** *tensor* (Baroni and Lenci 2010)

- target word,
  - <type of relation-occurrence>,
  - co-occurrence values
- 

Two textual models:

- **M1**: textual matrix with the 20 nominal subject and object co-occurrences with the highest LMI value;
- **M2**: textual matrix with all the nominal subject and object co-occurrences

# VISUAL RESOURCES

## Images from ImageNet (Deng et al. 2009)

### Kit fox, *Vulpes macrotis*

Small grey fox of southwestern United States; may be a subspecies of *Vulpes velox*

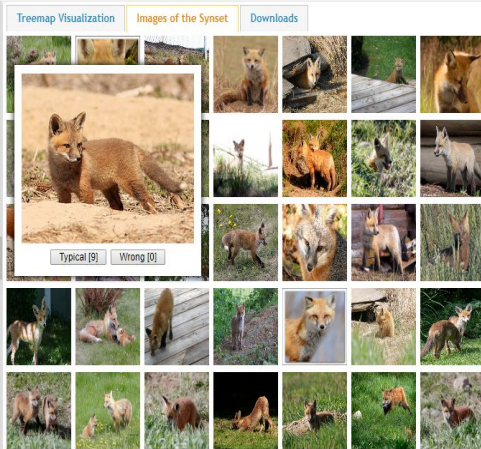
829  
pictures

60.83%  
Popularity  
Percentile

Wordnet  
IDs

Numbers in brackets: (the number of  
synsets in the subtree).

ImageNet 2011 Fall Release (32326)  
- plant, flora, plant life (4486)  
- geological formation, formation (17)  
- natural object (1112)  
- sport, athletics (176)  
- artifact, artefact (10504)  
- fungus (308)  
- person, individual, someone, some  
- animal, animate being, beast, brute  
- invertebrate (766)  
- homeotherm, homoiotherm, hor  
- work animal (4)  
- darter (0)  
- survivor (0)  
- range animal (0)  
- creepy-crawly (0)  
- domestic animal, domesticated  
- molar, moulter (0)  
- varmint, varment (0)  
- mutant (0)  
- critter (0)  
- game (47)  
- young, offspring (45)  
- poikilotherm, ectotherm (0)  
- herbivore (0)  
- peeper (0)  
- pest (1)



\*Images of children synsets are not included. All images shown are thumbnails. Images may be subject to copyright.

- Images are organized according to the semantic hierarchy of WordNet;
- label average accuracy: 99.7%;
- high resolution, open access images.

# VISUAL RESOURCES

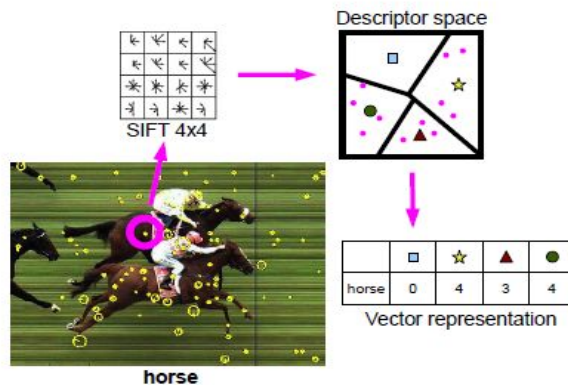
From the nouns co-occurrences selected for the M1 matrix



We extract visual representations and compute BoVW using **MMFeat** (Kiel, 2016)



**Scale Invariant Feature Transform** algorithm (Lowe, 1999) to find out image descriptors



: (

Some abstract names (*theory*, *experience*, *etc.*) do not have a visual representation available (Hill and Korhonen 2014, Anderson 2017).



: )

For all the other concepts, Visual Words representing the same concept are aggregated.

>> We calculate VW centroid  
>> Using these values we built the visual matrix **MV**

# BUILDING MATRICES

The data collected are then organized in matrices.

In textual matrices: verbs in rows, nouns in columns

In the visual matrix: verbs in rows, nouns representations in columns, and images' centroids in entries.

$$\begin{matrix} & c_1 & c_2 & \dots & c_n \\ \begin{matrix} t_1 \\ t_2 \\ \vdots \\ t_m \end{matrix} & \begin{pmatrix} w_{t_1, c_1} & w_{t_1, c_2} & \vdots & w_{t_1, c_n} \\ w_{t_2, c_1} & w_{t_2, c_2} & \vdots & w_{t_2, c_n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{t_m, c_1} & w_{t_m, c_2} & \vdots & w_{t_m, c_n} \end{pmatrix} \end{matrix}$$

Matrices were reduced applying SVD with 300 latent dimension:



$$A = U\Sigma V^t$$

PART 4-

PRELIMINARY RESULTS

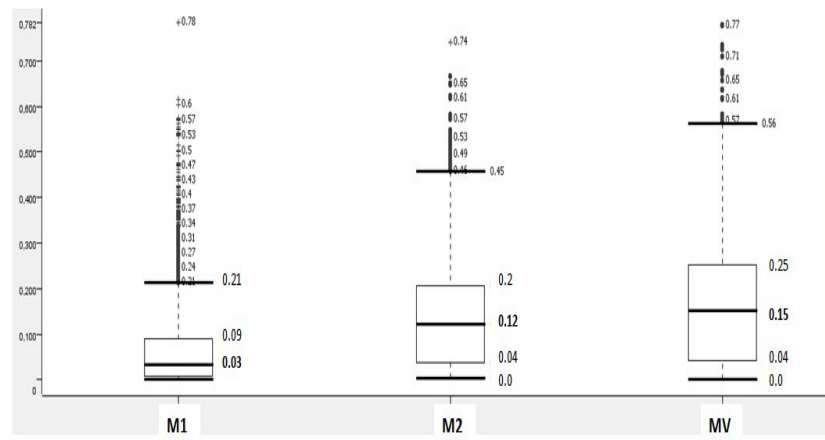


Similarity among verbs is evaluated computing the cosine between the two vectors within the distributional space.

MV has the highest values of the cosine similarity values

It is possible to define the semantic similarity between verbs using visual information

	M1	M2	MV
<b><i>decide-choose</i></b>	0.16	<b>0.3</b>	0.21
<b><i>pursue-realize</i></b>	0.06	0.04	<b>0.36</b>
<b><i>save-protect</i></b>	0.01	0.12	<b>0.14</b>

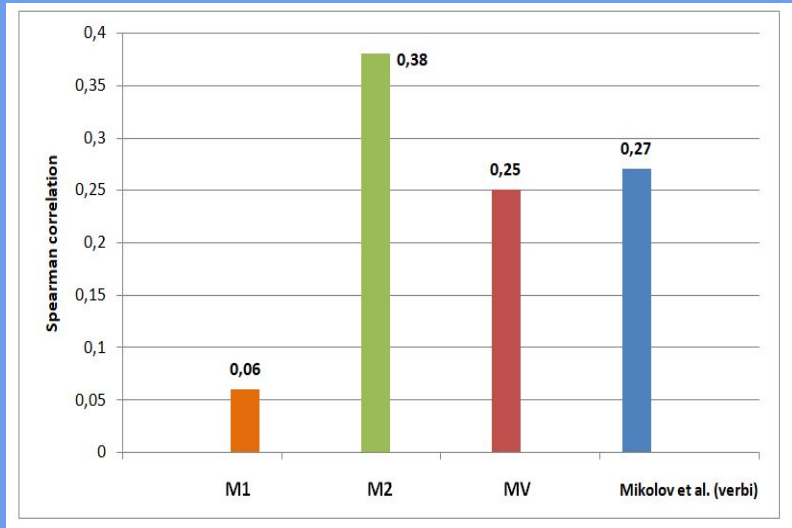


# EVALUATION WITH SIMLEX-999

This comparison presents three critical points:

1. Distributional models usually are not able to identify the similarity between two words regardless their association degree;
2. Within SimLex, antonyms are marked with a low similarity score;
3. DSMs perform worst in recognizing the similarity between verbs compared to other POS

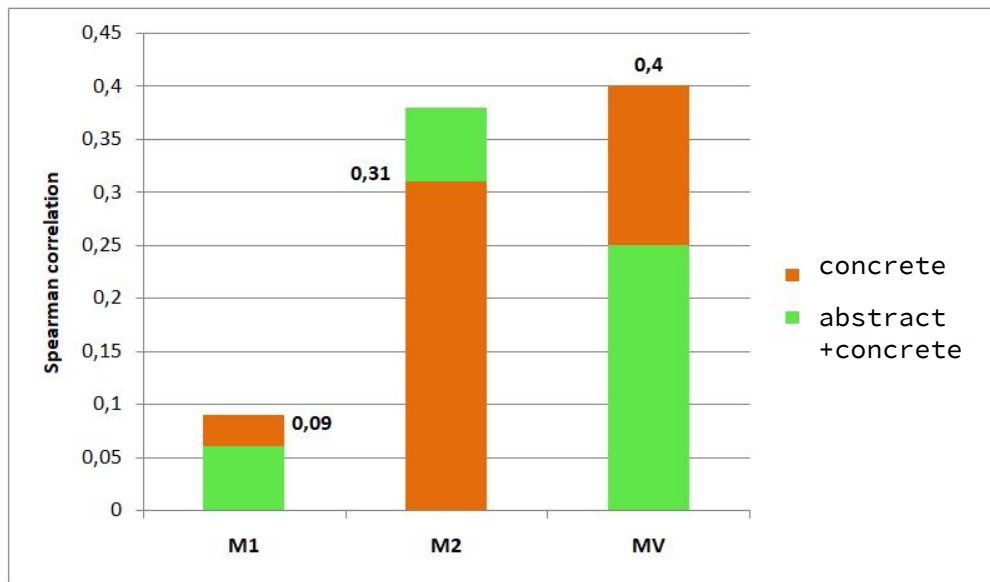
For the same number of co-occurrences considered, MV achieved a better result than M1 ( $p = 0.25$  vs  $p = 0.06$ )



**BUT!** The performance of textual model increases if the number of co-occurrences considered is increased.

MV is also competitive if compared to the model realized by Mikolov et al. (2013) ( $\rho = 0.27$ ) which uses the BoW to encode linguistic information related to verbs.

# WHAT IF WE CONSIDER JUST CONCRETE VERBS?

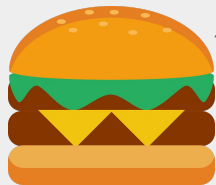


Substantial improvement in the performance of the visual model ( $\rho = 0.4$ )

...while little or no improvement was recorded for the textual models



Save - keep = 0.4



Save-preserve  
= 0.5



Save-rescue  
= 0.8

Aww man, how can I  
save my beloved?





Have you tried to  
defeat Browser?



Aww man, how can I  
save my beloved?





# PART 5-CONCLUSION

# CONCLUSION-1

The results obtained demonstrate that using perceptual information it is possible to effectively represent the semantic similarity between verbs.

- MV performs well when evaluated with SimLex
- For the same number of co-occurrences considered, MV performs better than M1
- MV is also competitive compared to other models (Mikolov et al. 2013)

# CONCLUSION-1

The results obtained demonstrate that using perceptual information it is possible to effectively represent the semantic similarity between verbs.

**BUT!** At the moment, visual and multimodal models produce no improvement in the representation of abstract concepts(Hill and Korhonen, 2014).

# CONCLUSION-2

STAY TUNED! :)

What about the future?

- Bigger is better (?) Towards a Multimodal Distributional Semantic Model.
- Using a new measure (image dispersion) to improve the ability of the model to learn and represent word meanings (especially abstract word meanings).

# CONCLUSION 3 - WHY

Possibility to realize a global computational theory that describe the mechanism underlying human learnings

## IS THIS IMPORTANT?

For the improvement of systems which recognize, classify and describe images:



- AI systems
- dialog models
- different systems and domains

# THANK YOU FOR YOUR ATTENTION



Want to know more? Contact me:  
Irene Sucameli  
[irene.sucameli@phd.unipi.it](mailto:irene.sucameli@phd.unipi.it)