

your title

Your Name

Department of Computer Science  
TALP Group

Advisors:  
Prof. John  
Dr. Rosa

PhD Thesis Defense  
Feb 10, 2021



# Table of Contents

Problem Identification

# Table of Contents

Problem Identification

# Highlight with a block

A Fixed lexicon [some text ].

B Lexicon free [some text ].

# Highlight with a block

A Fixed lexicon [some text].

B Lexicon free [some text].

## Add Block in order A ...

**A Lack of public dataset:** Most state-of-art deep .....

## Add Block in order A ...

**A Lack of public dataset:** Most state-of-art deep .....

**B Fast and easy to re-train:** Statistical Language ....

## Add Block in order A ...

**A Lack of public dataset:** Most state-of-art deep .....

**B Fast and easy to re-train:** Statistical Language ....

**C** The system can be used as a .....



## Add Block in order A ...

A **Lack of public dataset:** Most state-of-art deep .....

B **Fast and easy to re-train:** Statistical Language ....

C The system can be used as a .....

D This **hybrid approach** between deep learning .....

# Literature review block

Work addresses scene understanding, and benefit from combining text cue and visual context in image or text retrieval:

## Lexicon Generation

Patel et al. (2016)

generation of new lexicon with **topic modeling**

## Logo Retrieval

Karaoglu et al. (2017)

learn **textual information** from logos

## Text detection

Prasad et al. (2018)

using object information for text detection

# Literature review block

Work addresses scene understanding, and benefit from combining text cue and visual context in image or text retrieval:

Lexicon Generation

Patel et al. (2016)

generation of new lexicon with **topic modeling**

Logo Retrieval

Karaoglu et al. (2017)

learn **textual information** from logos

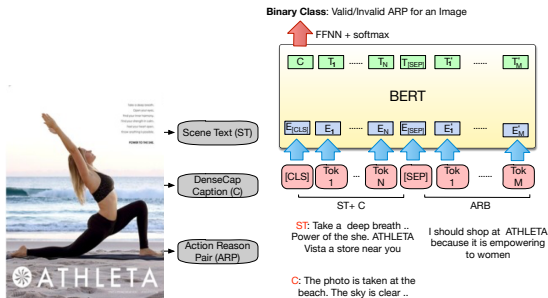
Text detection

Prasad et al. (2018)

using object information for text detection

# Image

- The task involves detecting the viewer's interpretation of an Ad image captured as text.
- Fine-tune BERT is used to learn textual and visual cues.
- Google Vision API is used to extract scene text information.



# Table

- The ULM is based on a combined corpus ....

Unique Count of Textual Data				
Dictionary	words	nouns	verb	adjectives
Dict-90K words level	87,629	20,146	6,956	15,534
Language model	8870,209	2695,906	139,385	824,581

# Fade out text

- The main limitation of this approach is that **depends on the baseline softmax** output to re-rank the most closely related word

As in the example, the semantic relatedness score **suppresses unrelated words** and **boosts the most probably related word** by simple dot product multiplication. (Visual context: parking meters)

$w$	Text Spotting Model		Visual Re-ranker Model
$w_1$	quotas	0.5	$5.4e-7$
$w_2$	quartos	0.1	$5.2e-8$
$w_3$	<b>quarters</b>	0.05	$9.0e-9$

## Attention 1

$$c_t = \sum_{j=1}^T \alpha_{tj} h_j, \alpha_{tj} = \frac{\exp(e_{tj})}{\sum_{k=1}^T \exp(e_{tk})}, e_{tj} = a(s_{t-1}, h_j)$$

## Attention 2

$$c_t = \sum_{j=1}^T \alpha_{tj} h_j, \alpha_{tj} = \frac{\exp(e_{tj})}{\sum_{k=1}^T \exp(e_{tk})}, e_{tj} = a(\textcolor{red}{s}_{t-1}, h_j)$$

## Figure

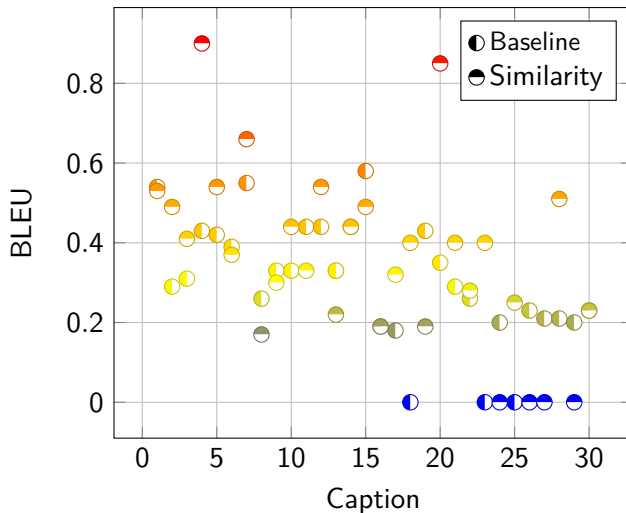




Figure 2

