



# Automatic Understanding of Image and Video Advertisements

Zaeem Hussain, Mingda Zhang, Xiaozhong Zhang, Keren Ye, Christopher Thomas, Zuha Agha, Nathan Ong, Adriana Kovashka  
University of Pittsburgh

IEEE 2017 Conference on  
Computer Vision and Pattern  
Recognition



## Introduction

- Advertisements *implicitly* persuade viewers to take certain actions.
- Understanding ads requires more than recognizing physical content.



Recognized Concepts (Clarifai):

Car, Street, Transportation System, Traffic, Road, City, Pavement, Crossing, ...

Image Caption (Vinyals et al.):

A red car driving down a street next to a traffic light.

True Meaning in Advertisement:

Automobile drivers should be cautious to avoid crashing into cyclists as they share the road.

- We propose the novel problem of **automatic advertisement understanding**, and provide two datasets with rich annotations.
- We analyze the common persuasive strategies: symbolism, atypical objects, physical processes, cultural knowledge, surprise/shock, etc.
- We present baseline experiment results for several prediction tasks.

## Dataset Collection

- Our dataset contains 64,832 image ads and 3,477 video ads, each annotated by 3-5 human workers from Amazon Mechanical Turk.

Image	Topic	204,340	Sentiment	102,340	Q+ A Pairs	202,090
	Symbol	64,131	Strategy	20,000	Slogan	11,130
Video	Topic	17,345	Sentiment	17,345	Q+ A Pairs	17,345
	Fun/Exciting	17,374	English?	15,380	Effectiveness	16,721

## Experiment Summary

Prediction on Images	Accuracy	Prediction on Videos	Accuracy
Topic	60.34%	Topic	35.1%
Sentiment	27.92%	Sentiment	32.8%
Question-Answering	11.96%	Question-Answering	8.83%
Symbolism	15.79%	Funny	78.6%
	(F-score)	Exciting	78.2%

## Dataset Overview

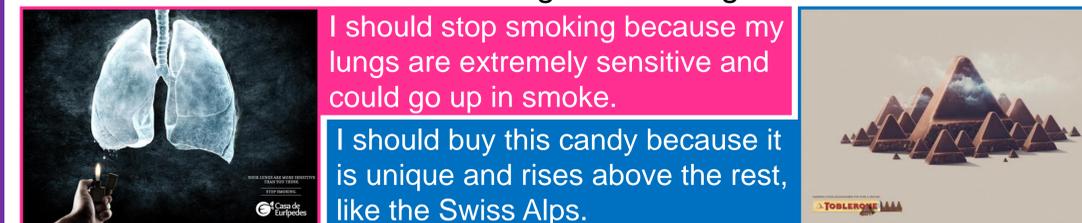
- 38 topics including commercials and public service announcements



- 30 sentiments indicating how ads emotionally impress viewers

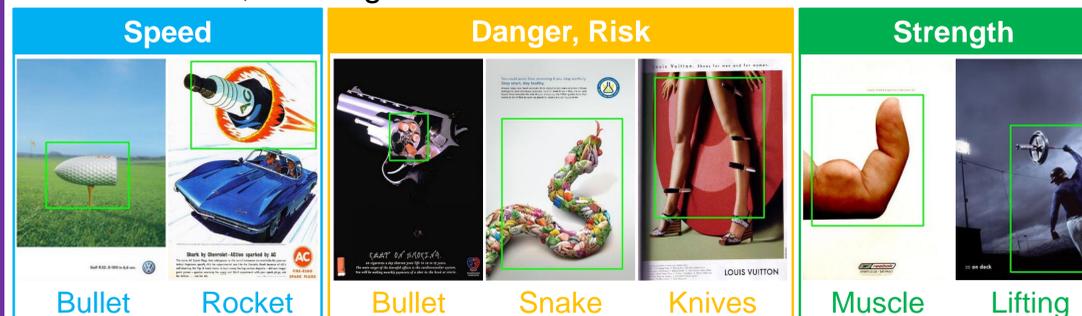


- Questions and answers revealing the messages behind the visual ads



## Symbolism Detection

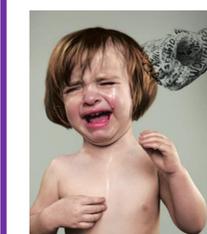
- Symbolism efficiently links physical content with abstract concepts.
- Decoding symbolism relies on outside knowledge and cultural associations, allowing viewers make reasonable inferences from ads.



- Attention model is used to detect symbolism in ads achieving **15.79% / 26.84%** F-score when distinguishing **221 / 53** common symbols.

## Answering Questions about Ads

- To decode the messages behind ads, vision systems should answer questions like **why should I take the action that the ad suggests?**



What should I do? I should be careful what words I use on my kid.  
Why? Because words can hurt as much as fists.



Q: Why should I be careful what words I use on my kid?  
A: Because words can hurt as much as fists.

- We test how well VQA models can answer such challenging questions
  - Two-layer LSTM is used for encoding the *question sentence*.
  - VGGNet is used for encoding the *image*.
  - *Full-sentence answers* are trimmed to most "contentful" *label* using TFIDF.
- The 1000-way classification baseline achieves **11.48%** accuracy.
- Using the probability distribution over *symbols* as an extra feature improves accuracy to **11.96%**; richer use of symbolism is future work.
- Example:



Why should I wear Vera Wang Perfume?  
Ground Truth: sexy, beautiful, attractive

Detected Symbols:  
sex, beauty, sexy, sex appeal, seduction, ...

QA Baseline : smell  
QA + Symbolism: sexy

## Dataset Release

- Image, video, annotations and statistics are all available from:  
<http://cs.pitt.edu/~kovashka/ads>

## Acknowledgement

- NSF CISE CRII (Award #1566270)  
*Automatically Understanding the Messages and Goals of Visual Media*
- Google Faculty Research Award