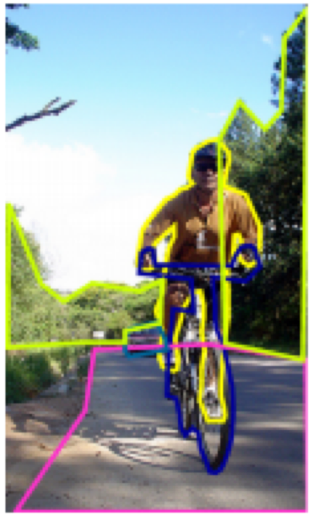
VISUAL DEPENDENCY PARSING AND VISUAL SENTIMENT ANALYSIS

Natalie Parde
parde@uic.edu

CS 594: Language and Vision
Spring 2019

What is visual dependency parsing?

- A method for extracting structured relational knowledge from images.

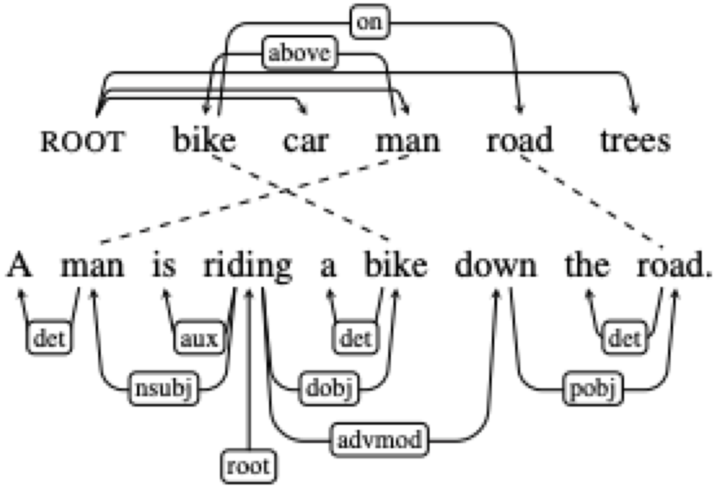


(a)

A man is riding a bike down the road.
A car and trees are in the background.

(b)

Desmond Elliott and Frank Keller. (2013). Image Description using Visual Dependency Relations. In *Proceedings of EMNLP*, http://anthology.aclweb.org/D/D13/D13-1128.pdf



(c)

Figure 1: (a) Image with regions marked up: BIKE, CAR, MAN, ROAD, TREES; (b) human-generated image description; (c) visual dependency representation expressing the relationships between MAN, BIKE, and ROAD aligned to the syntactic dependency parse of the first sentence in the human-generated description (b).

Visual dependency parsing is used to facilitate better image understanding.

Early Work:

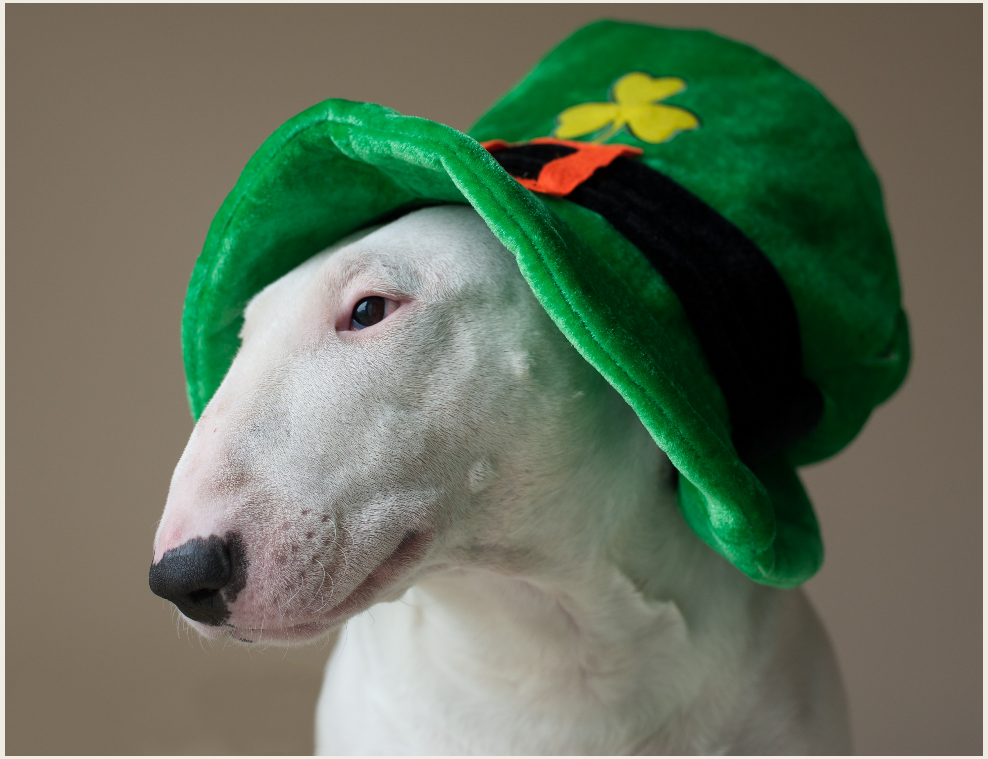
- **1961: Minsky suggests preliminary methods for describing patterns**
  - Minsky, M. (1961). Steps toward Artificial Intelligence. In *Computers and Thought*, Feigenbaum and Feldman (Eds.), pp. 466-450. McGraw-Hill, New York.
- **1964: Kirsch proposes that a generative grammar be used to do so**
  - Kirsch, R. A. (1964). Computer Interpretation of English Text and Picture Patterns. In *IEEE Transactions on Electronic Computers*, EC 13, pp. 363-376.
- **1971: Clowes introduces a rule-based linguistic pattern description language**
  - Clowes, M. (1971). On Seeing Things. In *Artificial Intelligence*, 2(1), pp. 79-116.

Later:

- **1995: Ralescu suggests that image understanding relates to verbal image description**
  - Ralescu, A. (1995). Image Understanding = Verbal Description of the Image Contents. In The Journal of the Japanese Society for Fuzzy Theory and Systems, 7(4), pp. 739-746.
- **2010: Bateman et al. propose a semantic ontology for linguistic spatial expressions**
  - Bateman, J. A., Hois, J., Ross, R. and Tenbrink, T. (2010). A Linguistic Ontology of Space for Natural Language Processing. In Artificial Intelligence, 174(14), pp. 1027-1071.
- **2014: Xu et al. propose a model for image interpretation based on semantic relations between image regions**
  - Xu, K., Ba, J. L., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., Zemel, R., Bengio, Y. (2015). Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. In *Proceedings of the 32nd International Conference on Machine Learning*, pp. 2048-2057.

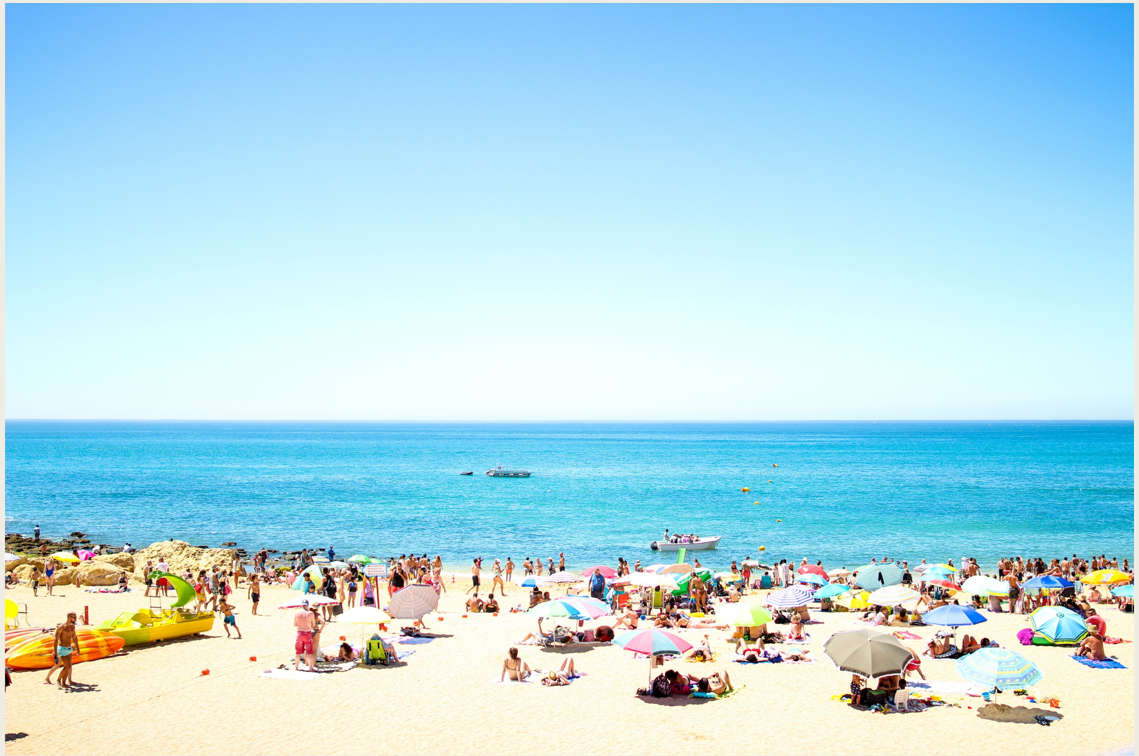
Why is it important to relate image components with one another?

Many image descriptions rely on relational assumptions.



☹ A green hat and a dog and a gray wall.

☺ A dog wearing a green hat in front of a grey wall.



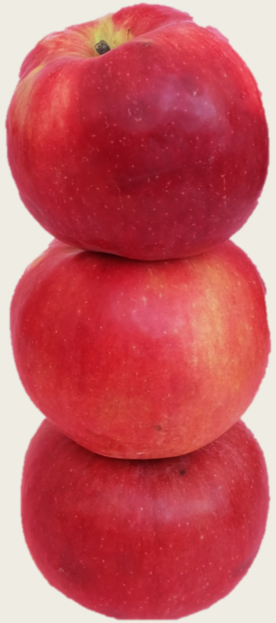
☹ Blue sky and water and sand and people.

☺ People on sand in front of water under a blue sky.

Relational knowledge can be crucial for situated dialogue.

Hand me the apple.

😐

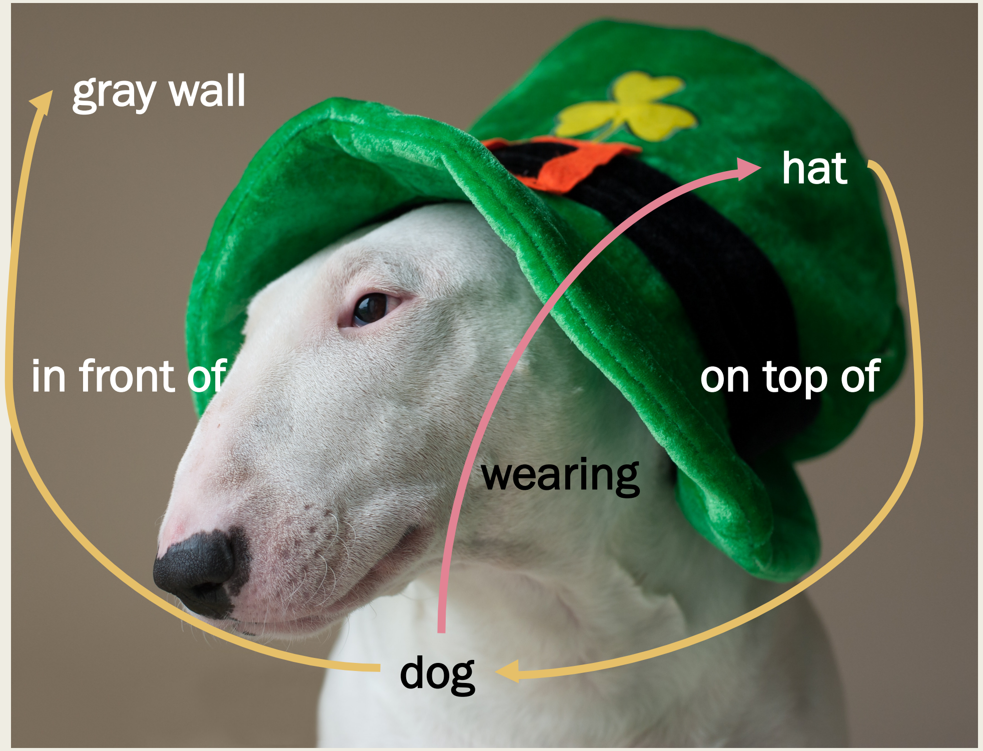


Hand me the apple on top of the other apples.

🙂

Three Steps in Visual Dependency Parsing

1. Identify image regions/components
   1. *Objects only?*
   2. *Object granularity?*
   3. *Background?*
2. Identify visual relationships between image components
3. Map visual relationships to semantic relationships

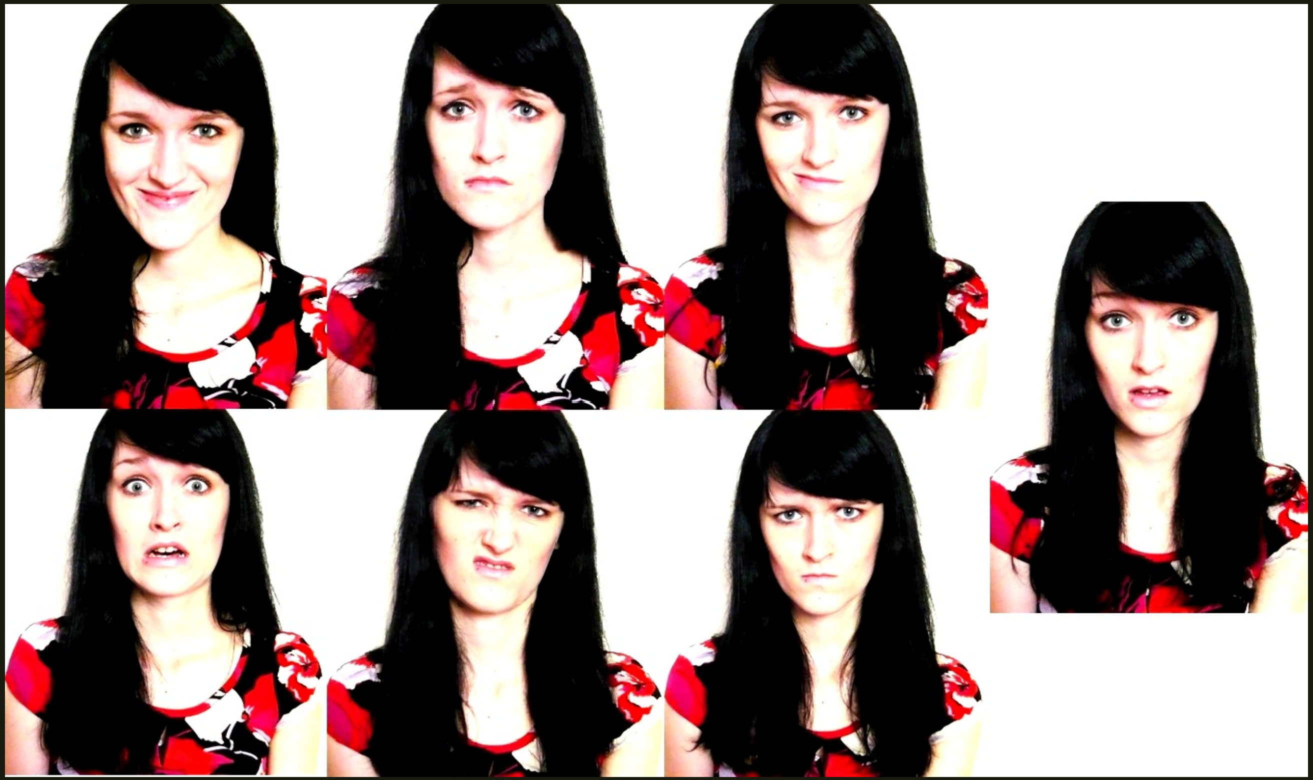


What are some ways that this can be done?

- Build visual dependency graphs between objects in training images, and learn mappings between those graphs and dependency parses of the images' descriptions
  - *Elliott, D. and Keller, F. (2013). Image Description using Visual Dependency Relations. In Proceedings of Empirical Methods in Natural Language Processing.*
- Build visual dependency graphs between objects in training images, and generate descriptions based on these graphs using a semantic grammar
  - *Lin, D., Fidler, S., Kong, C., and Urtasun, R. (2015). Generating multi-sentence natural language descriptions of indoor scenes. In Proceedings of British Machine Vision Conference.*

Resources for Visual Dependency Parsing

- Visual Dependency Parser
  - *https://github.com/elliottd/vdrparser*
  - *https://github.com/elliottd/vdrDataTools*
- Datasets
  - *Visual and Linguistic Treebank: http://homepages.inf.ed.ac.uk/s0128959/dataset/index.html*
  - *SentencesNYUv2: http://www.cs.toronto.edu/~fidler/projects/sentences3D.html*



😄 😀 🙂 😐 😕 🙁 ☹️

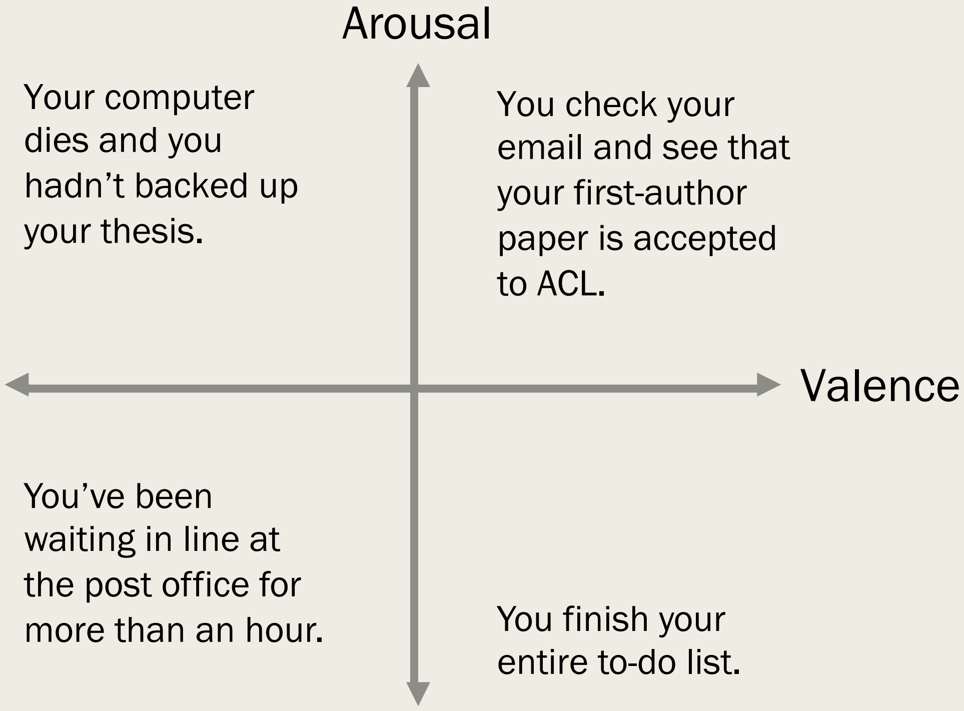
What is visual sentiment analysis?

- Visual sentiment analysis is a form of **affect recognition.**
- Affect recognition: Automatically determining a person's emotional state
  - *Happiness, sadness, fear, anger, surprise, disgust*
- Sentiment: Positive, negative, or neutral

Work in this area falls under the broader umbrella of affective computing.

- Affective computing: The study and design of systems capable of interpreting, processing, recognizing, and/or stimulating human affective states
- Highly interdisciplinary!
  - *Computer Science*
  - *Psychology*
  - *Cognitive Science*
- Widely considered to have been established by Picard in 1995
  - *Picard, Rosalind W. Affective computing. MIT press, 2000.*

Some Scales Used for Sentiment Analysis and Affect Recognition

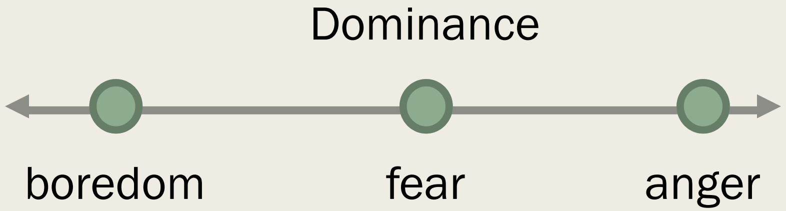


Your computer dies and you hadn’t backed up your thesis.

You check your email and see that your first-author paper is accepted to ACL.

You’ve been waiting in line at the post office for more than an hour.

You finish your entire to-do list.



Basic Emotions

- First introduced by Ekman[1]
- Can be viewed categorically or along a continuum

anger

disgust

happiness

fear

sadness

surprise

[1] Ekman, Paul, and Wallace V. Friesen. "Constants across cultures in the face and emotion." *Journal of personality and social psychology* 17.2 (1971): 124.

Many methods for affect recognition rely on data from a single modality.

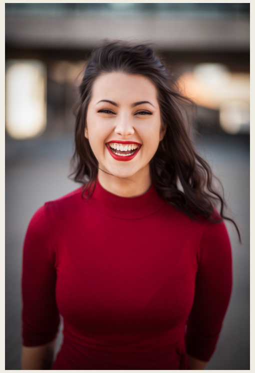
Text

- *Product reviews*
- *Tweets*
- *Narrative*

Images

- *Photos of people exhibiting sentiment/emotion*
- *Frames from video product reviews*

However, these methods are missing out on valuable real-world cues!



Just got the email …my work was nominated for best paper at AAAI!

Just got the email …my work was rejected from AAAI.

Multimodal affective computing leverages data from multiple input modalities.

- Common combinations:
  - *Text + Audio*
  - *Text + Audio + Video*
  - *Audio + Video*
- Less common but of growing interest:
  - *Text/audio/video + one or more of:*
    - EEG
    - EKG
    - Other data from wearable sensors
    - Typing behaviors

Two Types of Data

Natural: The labeled instance of emotion/sentiment occurred naturally, without any prompting.

Induced: The labeled instance was artificially induced. Emotions may be induced either consciously (e.g., by acting) or unconsciously (e.g., through predetermined external stimuli).

Pros and cons of each?

Characteristics to look for when analyzing affect.

Text	*Semantic Content* *Phrase Structure* *Complexity*
Image	*Facial expression* *Gesture*
Audio	*Volume* *Tone* *Pitch* *Speed*

Applications of Multimodal Affective Computing

- Mental health monitoring and assessment
- Measuring engagement
- Mood forecasting
- Image and video understanding
- Autism therapy
- Social Robotics

Resources for Multimodal Affective Computing

Recent Publication Venues

- Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media: https://peopleswksh.github.io/
- Workshop on Human Multimodal Language: http://multicomp.cs.cmu.edu/acl2018multimodalchallenge/
- Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis: https://wt-public.emm4u.eu/wassa2018/

Datasets

- CMU Multimodal Opinion Sentiment and Emotion Intensity Dataset: http://multicomp.cs.cmu.edu/resources/cmu-mosei-dataset/
- Tumblr Post Dataset: https://github.com/anthonyhu/tumblr-emotions

Lectures

- Technology and Emotions, by Rosalind Picard: https://youtu.be/ujxriwApPP4
- Multimodal AI: Understanding Human Communication Dynamics: https://youtu.be/dx2iwFEBmDU

Wrapping up….

- Overview of visual dependency parsing
- History of relational image understanding
- Steps involved in visual dependency parsing
- Resources for visual dependency parsing
- Overview of multimodal affective computing
- Scales used for sentiment analysis and affect recognition
- Types of data used for multimodal affective computing
- Applications of multimodal affective computing
- Resources for multimodal affective computing