A System for Image Understanding using Sensemaking and Narrative

Zev Battad, Mei Si

Department of Cognitive Science

Rensselaer Polytechnic Institute

Presented at Advances in Cognitive Systems 2021

11/18/2021

Introduction

- When explaining a set of images, human beings include a lot of information beyond what they see to tie things together.

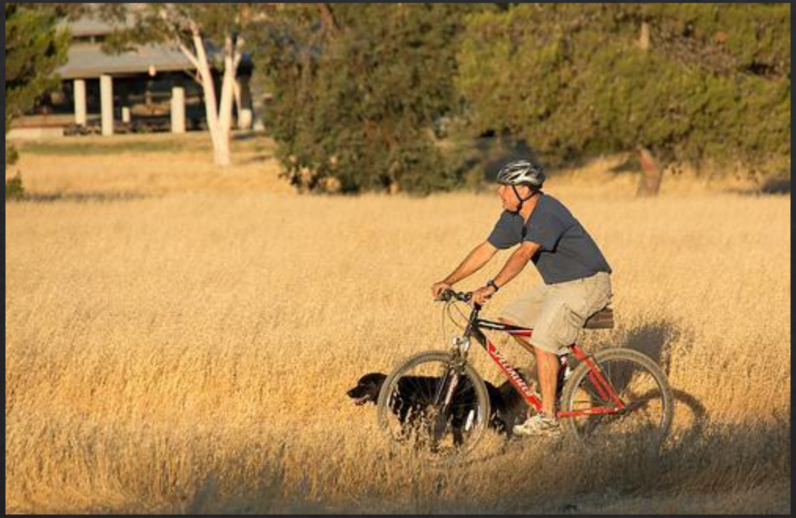
"Jerry was riding through a field with his bike on the way to get some food."

"While he was in the restaurant, a few dogs came up to his bike."

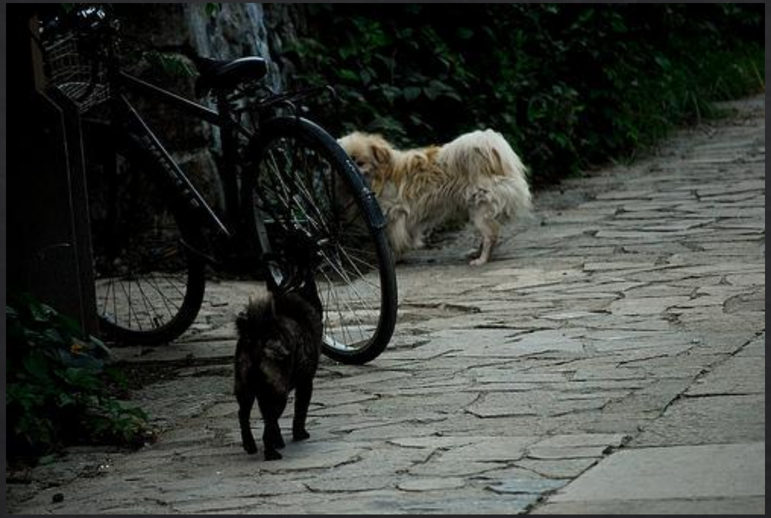
"Jerry thought they were scary so he called his friend Gertrude to help him scare the dogs away with her dog."

Introduction

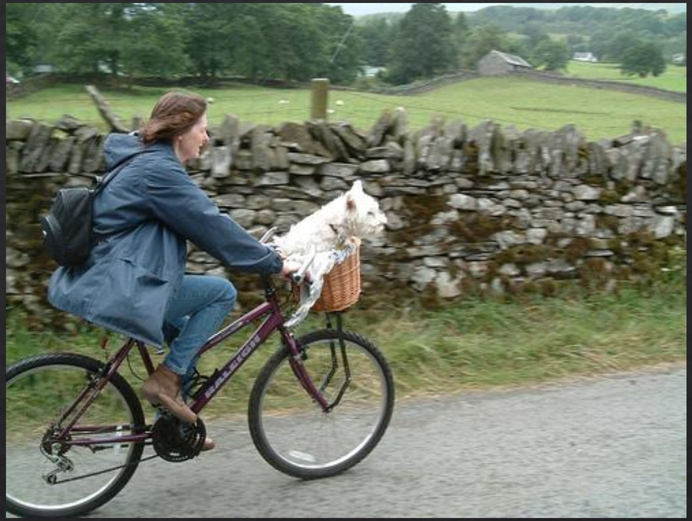
- When explaining a set of images, human beings include a lot of information beyond what they see to tie things together.



black dog in field, man has shadow, shadow in grass, man riding bike, man wearing helmet, dog beside man, dog has tongue, bike has front tire



short tail on black dog, tire on bicycle, spokes on wheel, four legs on dog, bike leaning on wall, dog next to bike, ground next to dog, tire on bike



rock on wall, letter on bike, woman riding a bike, backpack on woman, dog riding in the basket, stone fence along the rode

Introduction

Our goal is to create a computational sensemaking system that can create a knowledge graph from a sequence of images to support storytelling beyond the directly observable.

- The system's task is to interpret what is going on in the images.
- To accomplish this, it will need to introduce new entities, events, and evaluations, as well as connections between both new and old information.
- The information introduced has to be relevant and non-conflicting (e.g. it has to 'make sense').

Introduction

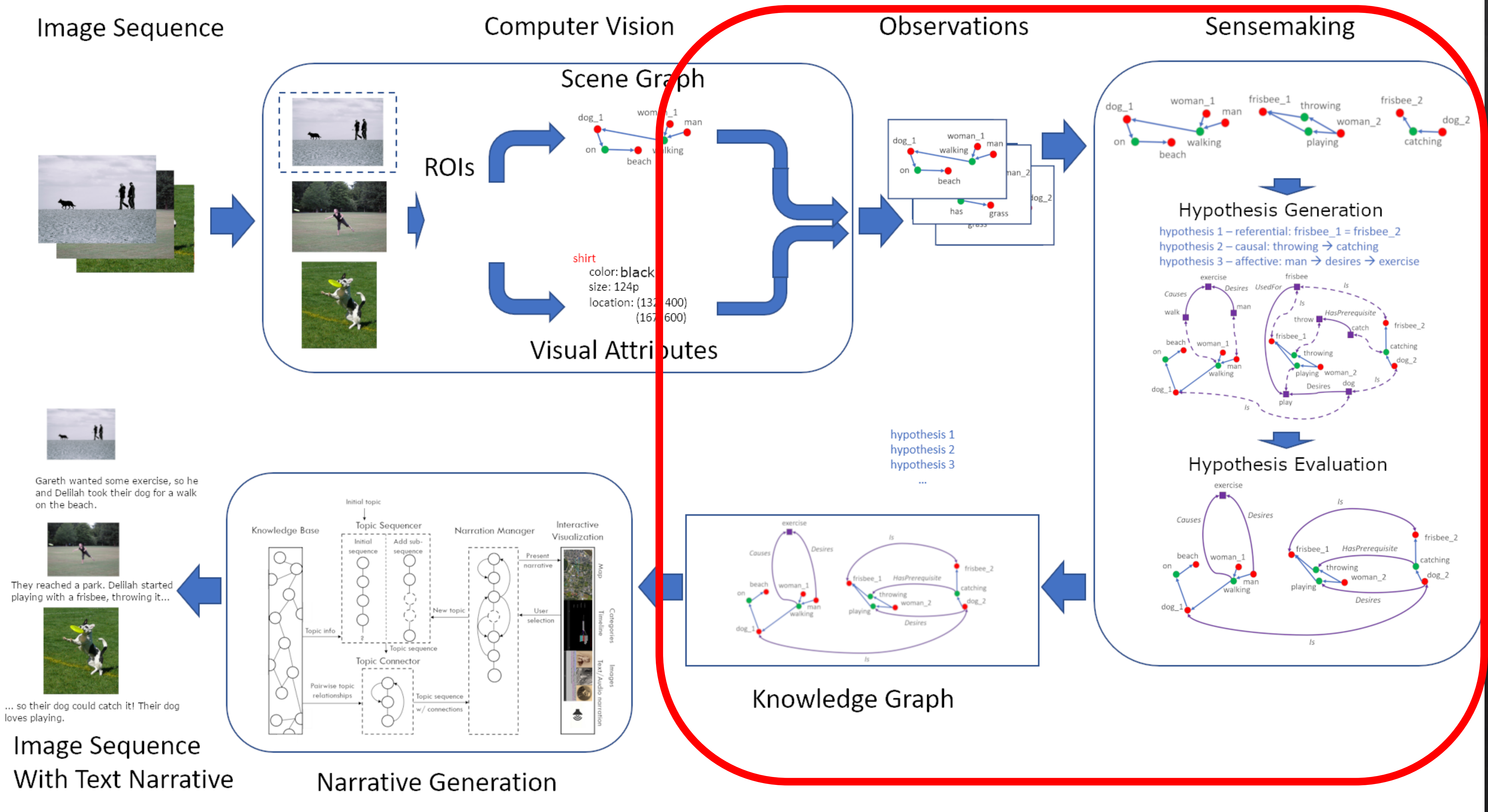
Our goal is to create a computational sensemaking system that can create a knowledge graph from a sequence of images to support storytelling beyond the directly observable.

- Bring in additional information.
- Establish connections.
- Make sure those connections and information are consistent.

Sensemaking - Connections

- What sort of connections do we mean?
- Look at categorizations of narrative connections in personal narratives.
- Spatial orientation, temporal orientation, topical consistency (Reese, 2011).
  - Topical consistency encompasses: causal links, personal emotional and motivational evaluations, and connections to previous stories or events.
- Spatial, temporal, referential continuity (Givon, 1992; Gernsbacher, 1995).
- Causal consistency (including causal links between events), character consistency (including character motivations) (Rideout, 2013)
- Spatial, Temporal, Causal, Referential, Affective (emotional/motivational)

Approach - Architecture



hypothesis 1 – referential: frisbee_1 = frisbee_2

hypothesis 2 – causal: throwing → catching

hypothesis 3 – affective: man → desires → exercise

hypothesis 1
hypothesis 2
hypothesis 3
...

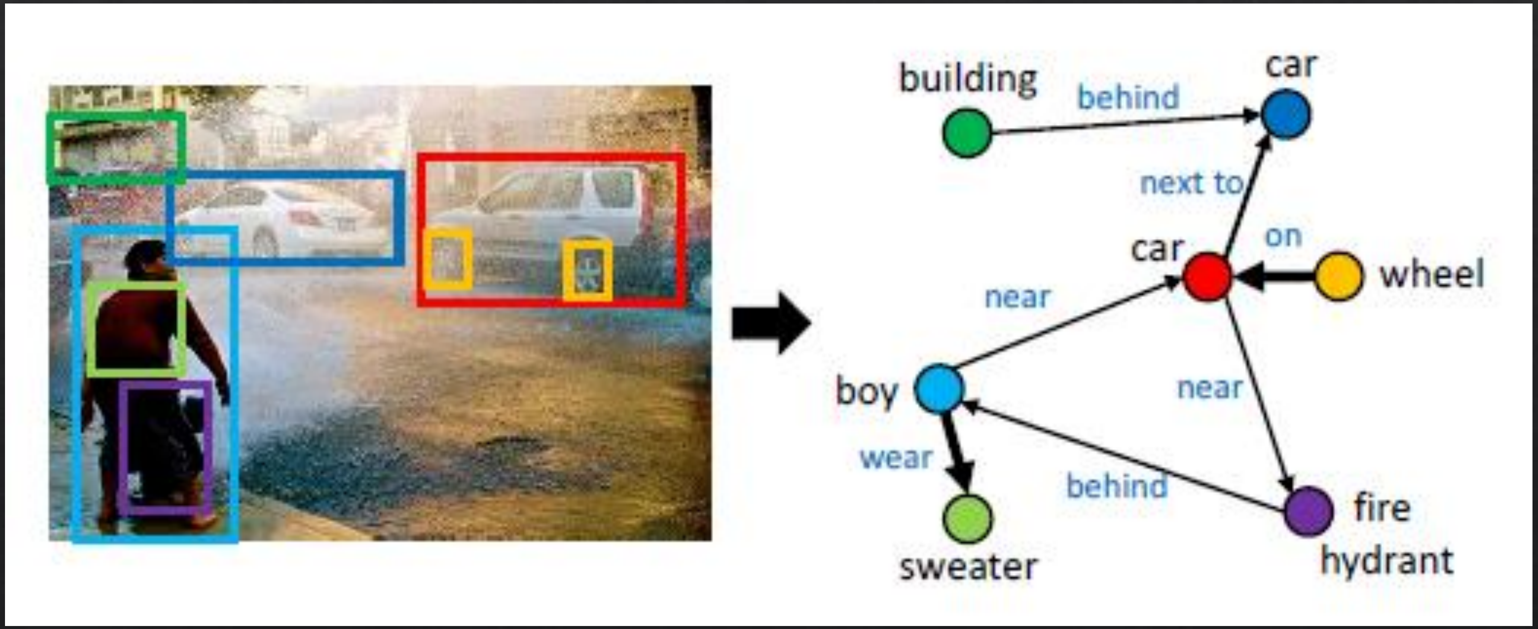
Gareth wanted some exercise, so he and Delilah took their dog for a walk on the beach.

They reached a park. Delilah started playing with a frisbee, throwing it...

... so their dog could catch it! Their dog loves playing.

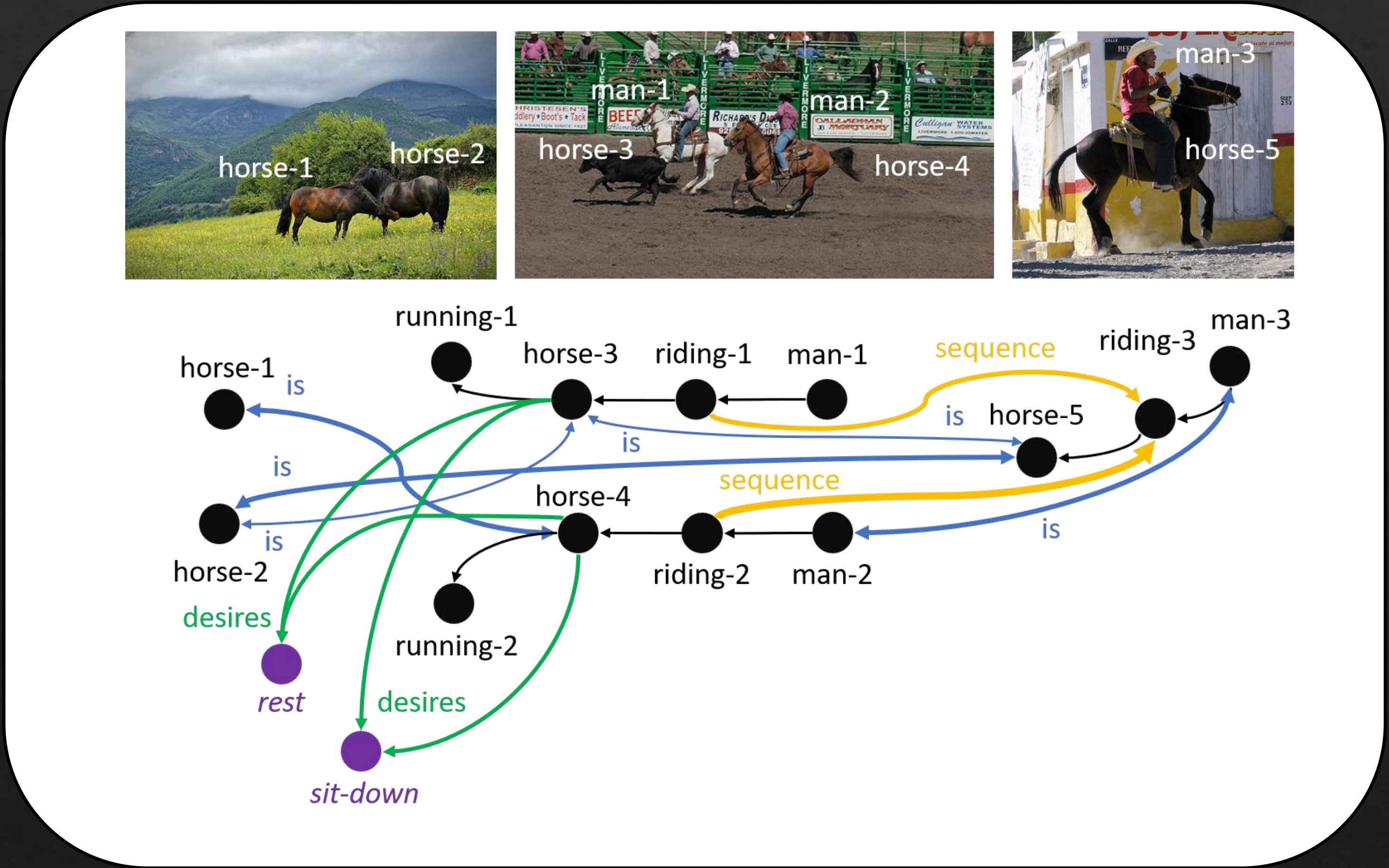
Approach – Visual Observations

- In lieu of a Computer Vision system, we use scene graphs annotations from the Visual Genome Dataset (Krishna et al., 2016).
- ~100,000 images with human-annotated scene graphs.



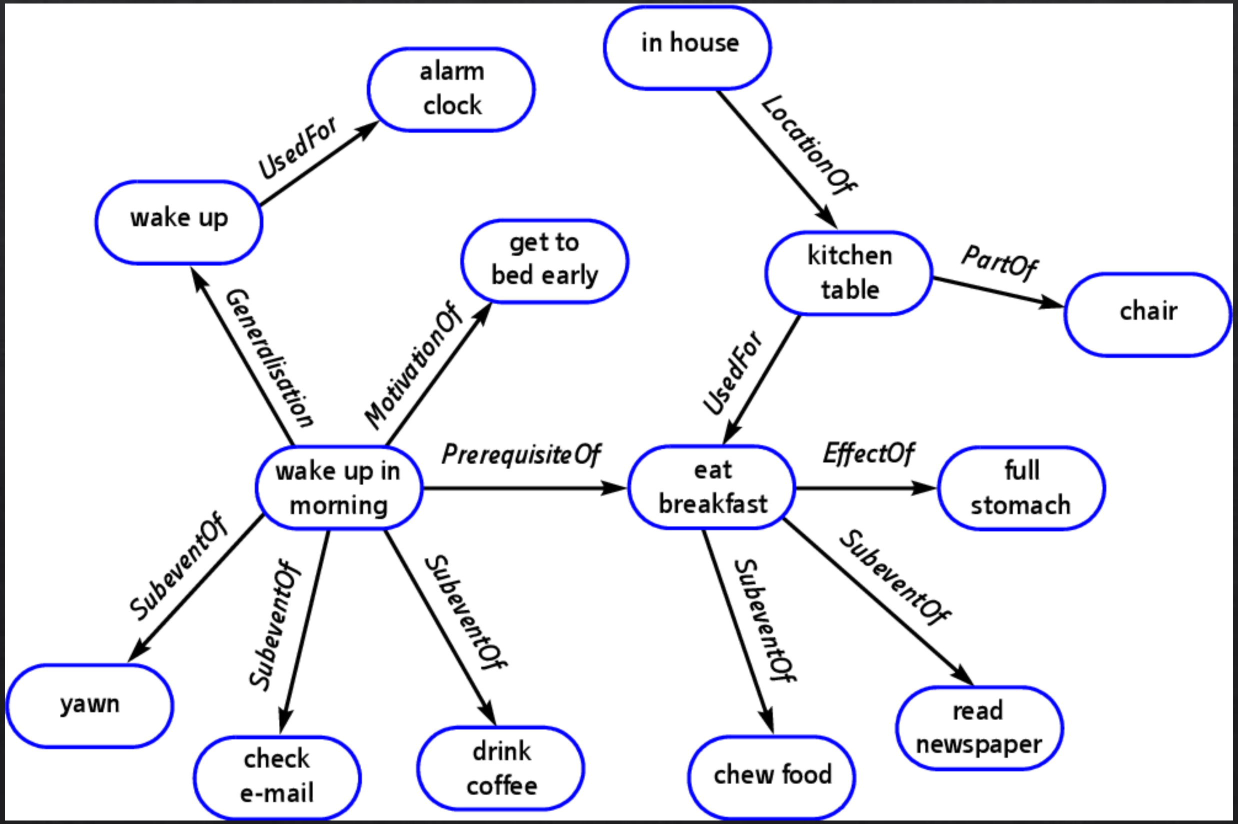
Approach – Sensemaking Subsystem

- From visual observations, produces a knowledge graph with hypothesized new information based on narrative connections.
- Hypothesis generation step – hypothesize narrative connections using external knowledge.
- Hypothesis evaluation step - choose which hypotheses to accept and which to reject for consistency.



Approach – Hypothesis Generation

- Generate hypotheses from commonsense knowledge source about where and whether narrative connections exist.
  - Spatial, Temporal, Causal, Referential, and Affective (emotional/motivational).
- ConceptNet (Speer, Chin & Havasi, 2017)



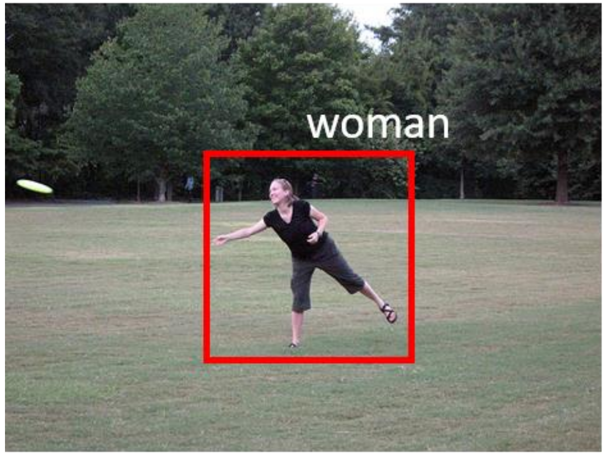
Approach – Hypothesis Generation

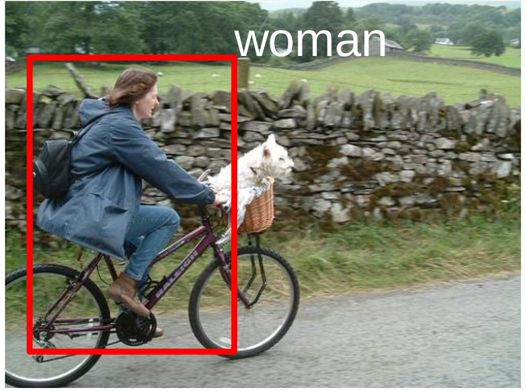
- Implemented hypothesis generation for Referential, Causal, and Affective relationships.
    - Interconnection of information.
        - Referential connects people, places, objects.
        - Causal connects events.
    - Introduction of new information.
        - Causal chains from external knowledge
        - Affective evaluations from external knowledge.
- Each draws from at least one of three different types of evidence
    - Observational evidence
    - Knowledge evidence
    - Prior hypotheses

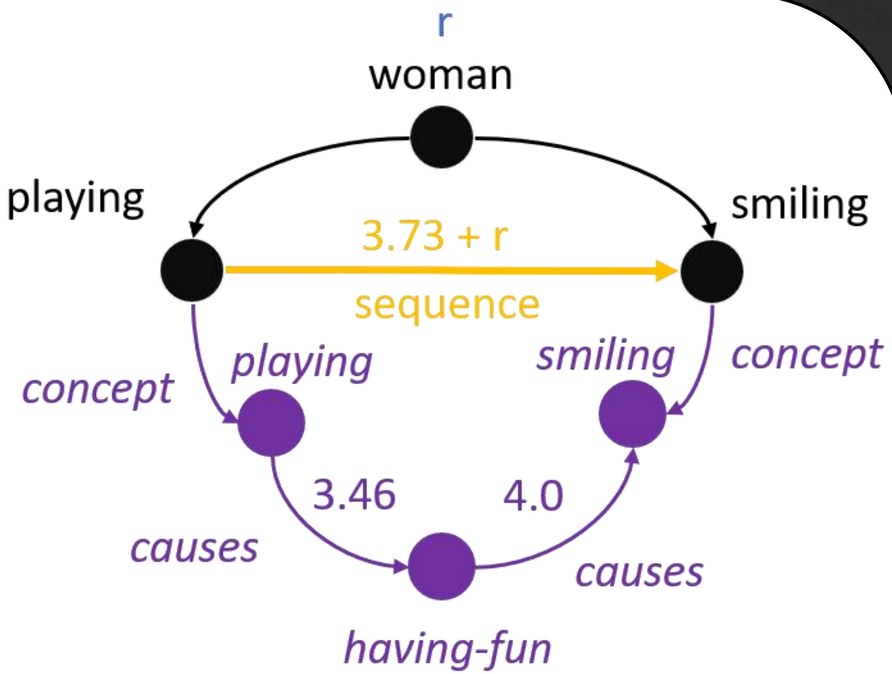
Approach – Hypothesis Generation

Evidence:

- Knowledge (causal path)
- Prior hypotheses (actors with referential hypothesis)







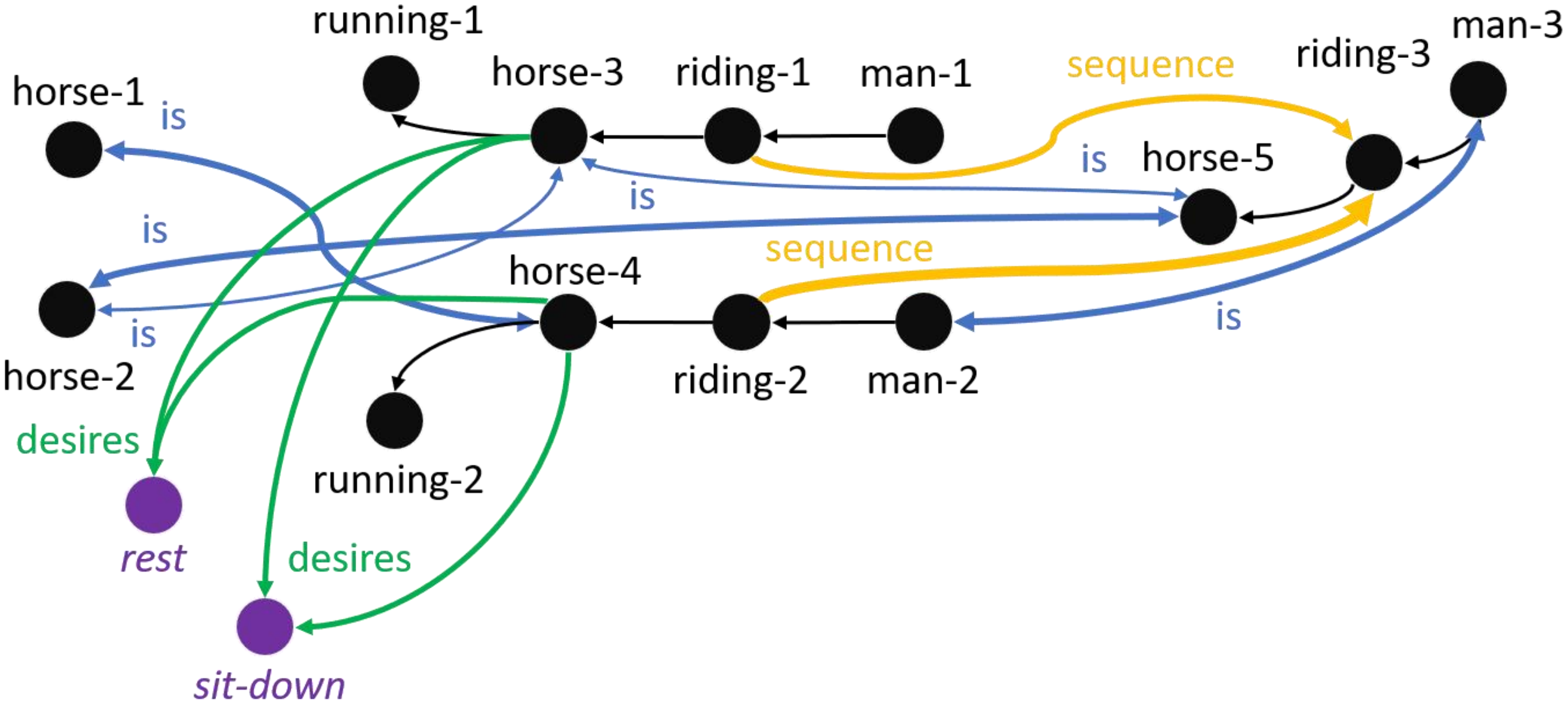
Approach – Hypothesis Evaluation

- Decide which hypotheses to accept and which to reject.
- Want to connect as much information as possible while weighing strength of hypotheses and avoiding inconsistencies.
- Multi-Objective Optimization Problem

Find the hypothesis set h_m that optimizes each objective functions:

- Connectivity k of graph G if all hypotheses in h_m are accepted.
- Density d of graph G if all hypotheses in h_m are accepted.
- Sum of strength of the score $s_{i,j}$ of each piece of evidence j for each hypothesis i in h_m
- Set of hypotheses must satisfy each constraint c_i to be considered feasible.
- Implemented two constraints on Referential hypotheses:
  - Non-redundancy (c_1): A referential hypothesis cannot assert that the same existent appears in the same image in two different places.
  - Transitive *is* assignment (c_2): If it is hypothesized that existent *a* and existent *b* are both existent *c*, then existent *a* must also be hypothesized to be existent *b*.

Approach – Sensemaking Subsystem



Evaluation

- Does the system lead to stories that are sensible to humans?
  - Would human beings rate the system’s output as plausible or sensible?
  - Do elements of the system’s output match elements of human-written stories?
- Does the system’s output confer some sort of cognitive benefit?
  - Does having the system’s outputs make the images make more sense than seeing the images by themselves?
  - Do people have an easier time reaching conclusions or answering questions about the images?
- How does it compare with other systems, e.g. end-to-end learning systems which also generate stories?

Evaluation

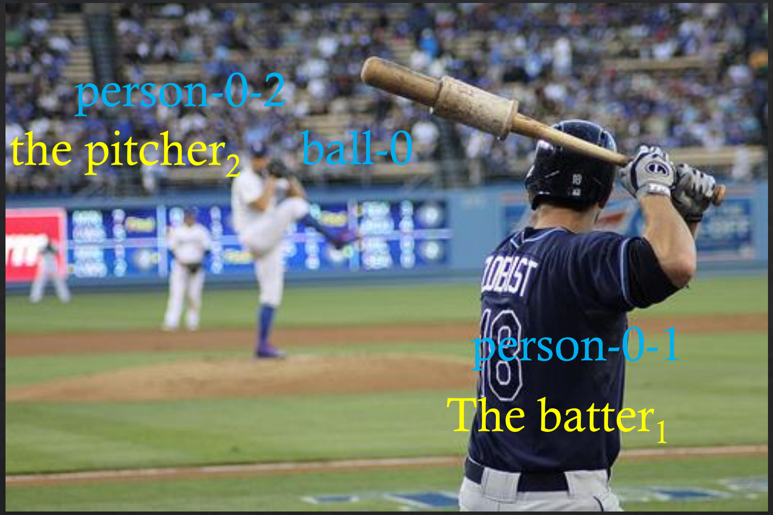
- Does the system lead to stories that are sensible to humans?
  - Would human beings rate the system’s output as plausible or sensible?
  - Do elements of the system’s output match elements of human-written stories?
- Does the system’s output confer some sort of cognitive benefit?
  - Does having the system’s outputs make the images make more sense than seeing the images by themselves?
  - Do people have an easier time reaching conclusions or answering questions about the images?
- How does it compare with other systems, e.g. end-to-end learning systems which also generate stories?

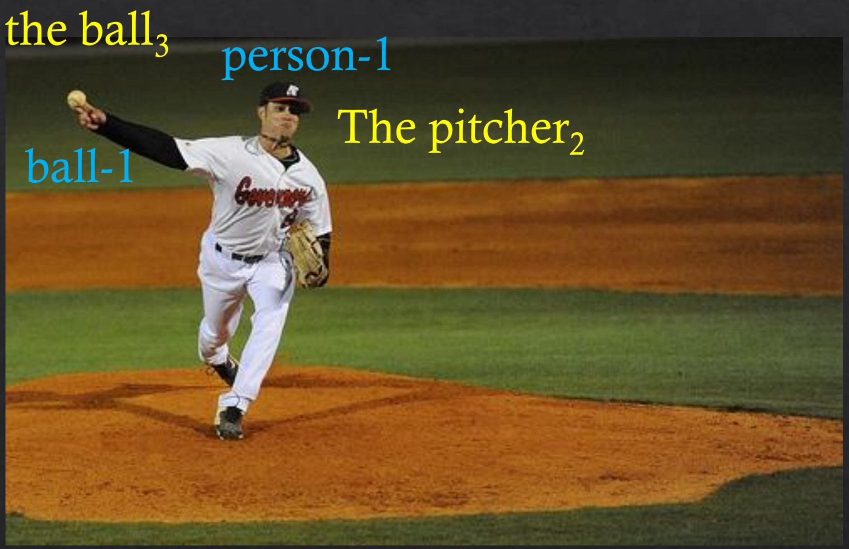
Evaluation

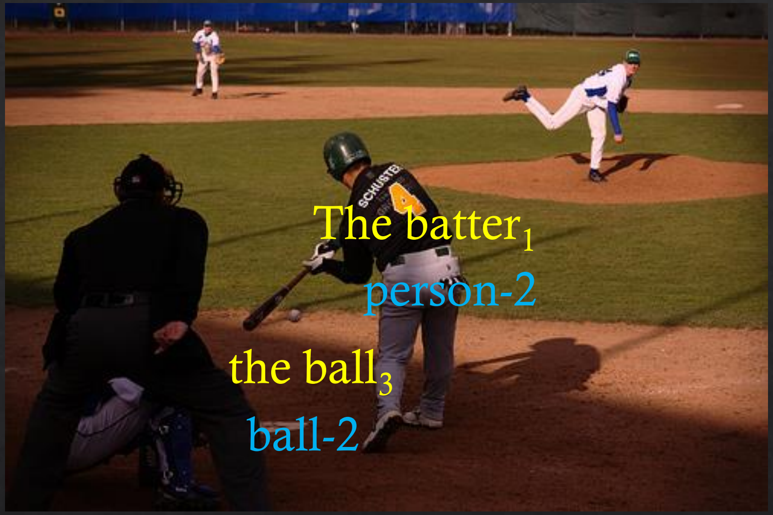
- Currently comparing system-generated knowledge graphs with elements of human-written stories.
- Suspension of disbelief for visual differences.
  - Sometimes human writers make referential connections between existents despite their visual differences (e.g. two different colored bicycles are the same).
  - How do these compare to referential hypotheses generated by the system?

Evaluation – Referential Hypothesis Comparison

Examining which coreference chains exactly match, partially match, and did not match system-generated referential hypothesis chains.







"[The batter]$_1$ was waiting for the pitch as [the pitcher]$_2$ wound up at the mound."

"Finally, it was the moment of truth. [The pitcher]$_2$ let [the ball]$_3$ go from his hand."

"[The batter]$_1$ swung the bat and there was a great crack as it struck [the ball]$_3$."

person-0-2 *is* person-1

ball-0 *is* ball-1 *is* ball-2

person-0-1 *is* person-2

Summary

- Building a computational sensemaking system to create knowledge graphs from sequences of images.
  - Support storytelling beyond the directly observable.
- Main effort is towards the sensemaking system.
  - Hypothesis generation through commonsense knowledge for narrative coherence connections (referential, causal, affective).
  - Hypothesis evaluation through multi-objective optimization to score, constrain, and select hypotheses.
  - Produces a knowledge graph with narrative connections.
- Evaluation comparing system-generated hypotheses with elements of human-written stories.
  - Suspension of disbelief for visual differences

Future Work

- Complete current evaluation of suspension of disbelief for visual differences.
- Compare non-observed information introduced by human writers with that introduced by the system's causal and affective hypotheses.
- Use analysis of above to modify current system.
  - Adjust objective score weights.
  - Add constraints.
  - Implement other relationships.
    - Temporal and Spatial.
- Expand evaluation
  - Ask humans to rate whether system's hypotheses are sensible.
  - Ask humans to answer questions/draw conclusions about images with or without the system's outputs.
  - Compare with intermediary outputs of other visual storytelling systems.
- Investigate human story phenomenon further.
  - What do people feel the need to tie together? Are there patterns as to what connections they employ and how?

Thank you!

Questions?