

AESOP: Abstract Encoding of Stories Objects and Pictures SICCVOCTOBER 11-17

Hareesh Ravi, Kushal Kafle, Scott Cohen, Jonathan Brandt, Mubbasir Kapadia {haravi, kkafle, scohen, jbrandt}@adobe.com, mubbasir.Kapadia@rutgers.edu



LIMITATIONS OF PRIOR WORK





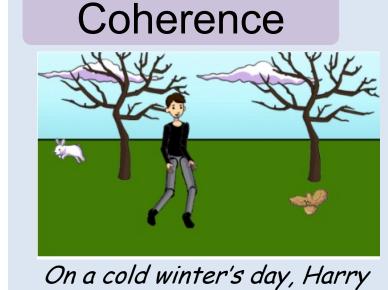
A cooking instruction video Two young girls are showing how to slice potatoes playing with lego to using a processor for french fries

- Factual information extraction [Graesser et. al. 1994]
- No temporal information in images. Video is aligned as a single entity. [Heider and Simmel 1994]
- Descriptive and Structured Text [Dunietz et. al. 2020]
- No Scene/Story creation



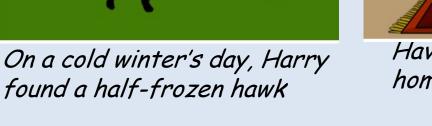
- - Incoherent Stories [Ravi et. al. 2018]
 - Models the perceptive process
 - Requires all of world knowledge

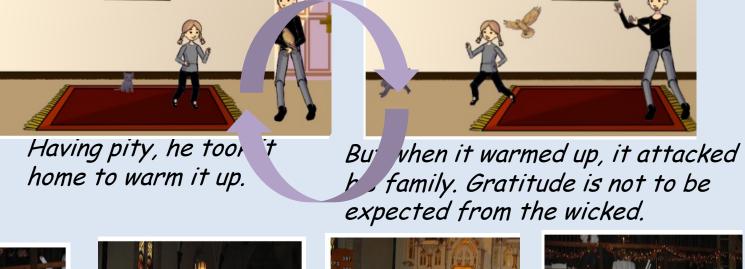
COMPARISON WITH EXISTING DATASETS



beautiful

day for him





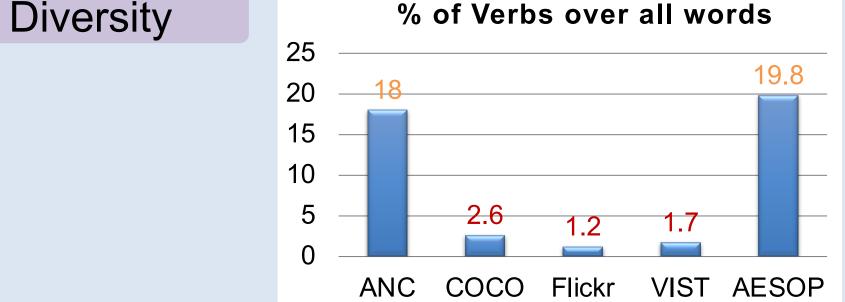


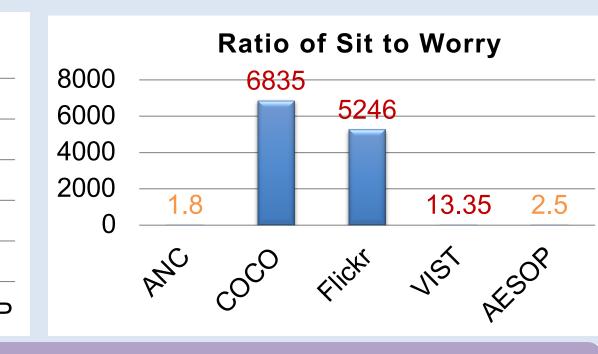


% of success

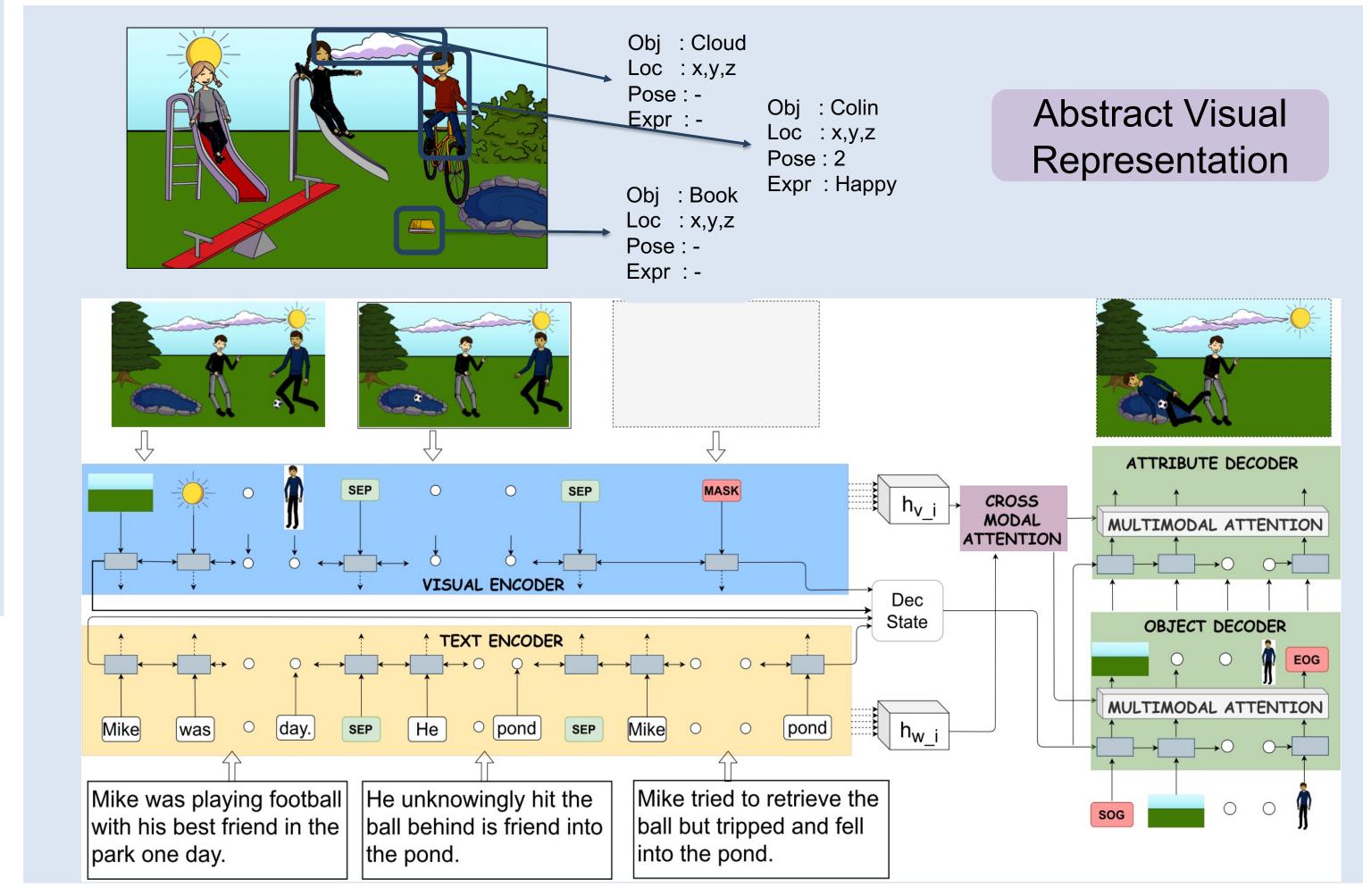
GT vs jumbled

97%





CO EVOLUTION OF VISUAL AND TEXTUAL



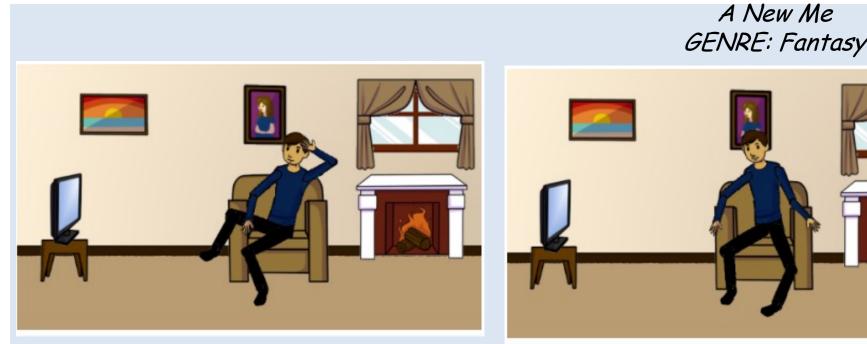
AESOP DATASET

Creativity over Perception [Antol et. . 2015]

Bounded Visual World



Causal and Coherent Narratives



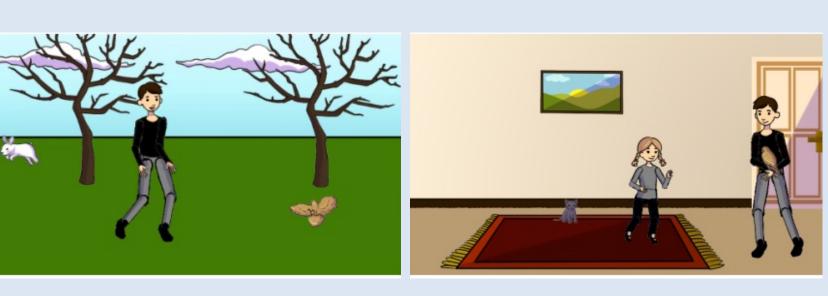
Mike was watching TV when suddenly an unusual promotion appeared. The advertisement was about buying your own clone.

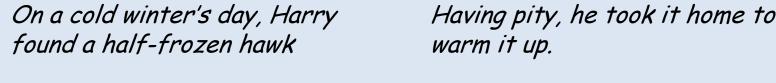
Mike was curious about the commercial, so he did what every bored man would do in his position; a bad decision.

And so he did, he called for a clone and there it was, another version of him. It was true; he thought it was a joke. Next time he would think twice about buying something on TV.

CREATION AND COMPREHENSION

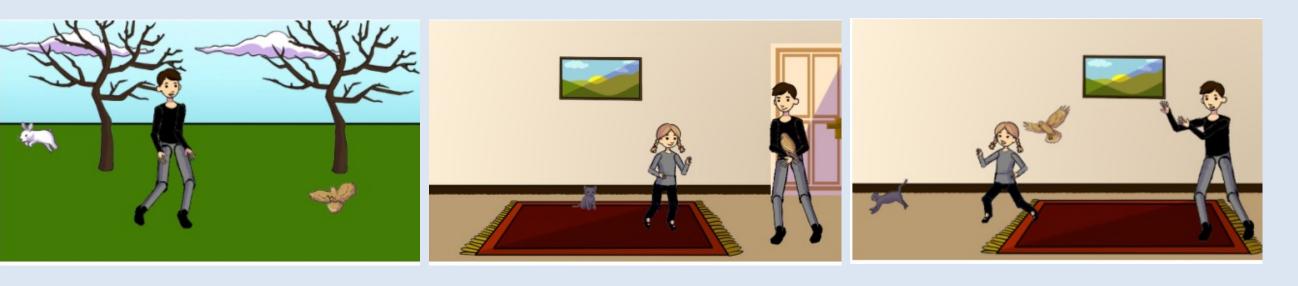
Assistant Illustrator



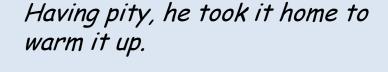


But when it warmed up, it attacked his family. Gratitude is not to be expected from the wicked.

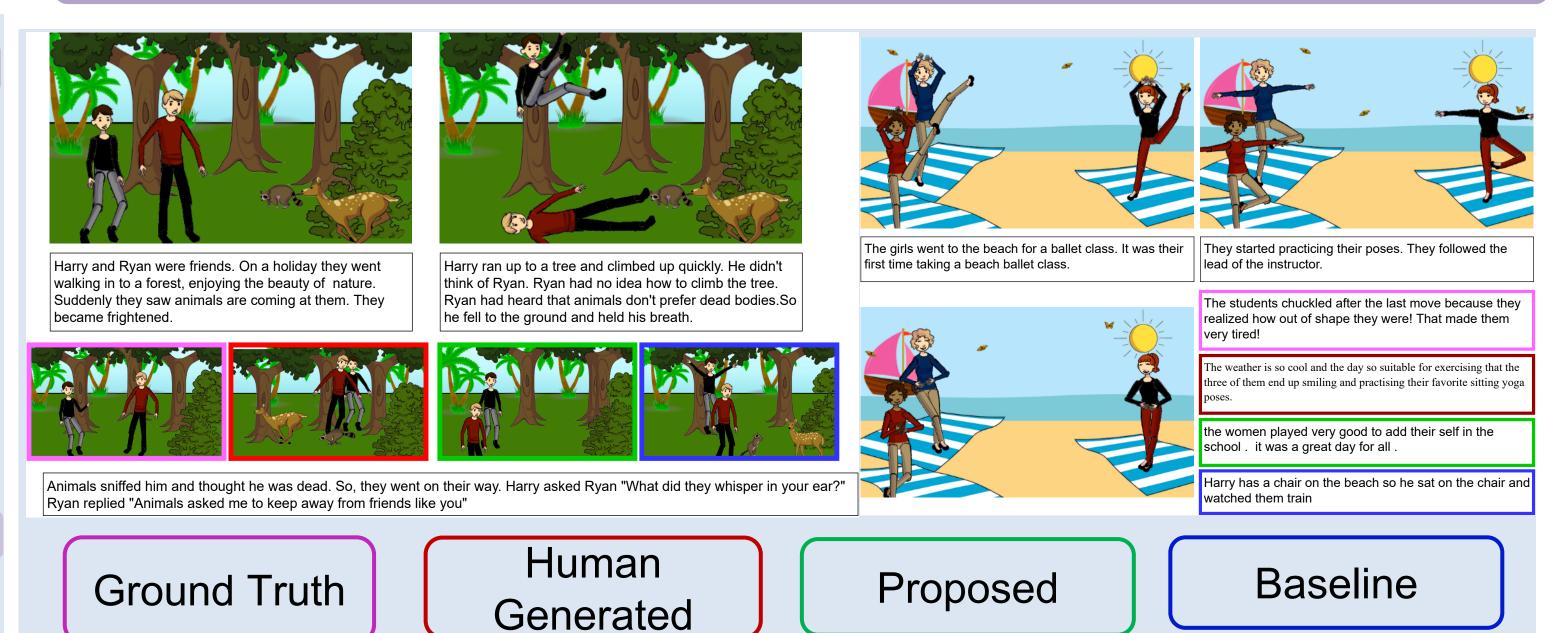
Assistant Writer



On a cold winter's day, Harry found a half-frozen hawk



RESULTS AND CONCLUSION



REFERENCES

- Constructing inferences during narrative text comprehension [Graesser et. al. 1994]
- An experimental study of apparent behavior [Heider and Simmel 1994]
- Show me a Story: Towards Coherent Neural Story Illustration [Ravi et. al. 2018]
- To test machine comprehension, start by defining comprehension [Dunietz et. al. 2020]
- VQA: Visual Question Answering [Antol et. al. ICCV 2015] • Cross-modal coherence modeling for caption generation [Alikhani et. al. ACL 2020]
- Visual Storytelling (VIST) [Hao et. al. NAACL 2016]

