Expressive Captioning: Textual Analysis

Lance Lebanoff, Jonathan Pham, David Hill
Mentor: Mahdi Kalayeh

UCF Center for Research in Computer Vision

National Science Foundation

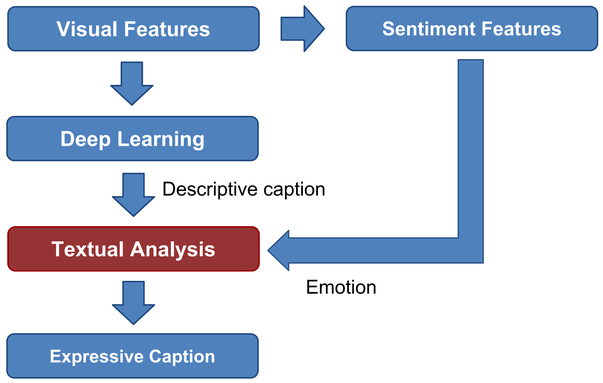
Problem Definition:

- Image captioning that exhibits the same emotion that is conveyed in the image.



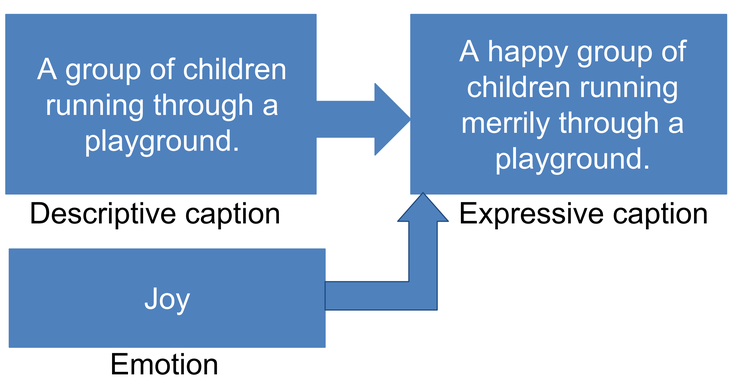
A happy group of children running merrily through a playground.

Pipeline:



Sub-Problem: Expressive Caption Conversion

- Descriptive caption: describes objects and actions displayed by the image
- Expressive caption: describes objects and actions, while expressing the same emotion conveyed in the image
- Goal: convert descriptive caption to expressive caption



Challenges:

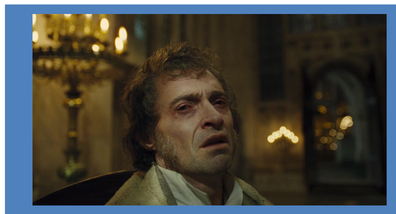
- No data set for descriptive sentences and corresponding expressive sentences
- No data set for expressive sentences and emotions conveyed by the sentences

Data Set Used:

- MPII Movie Description dataset [1]: movie clips and audio descriptions for blind people



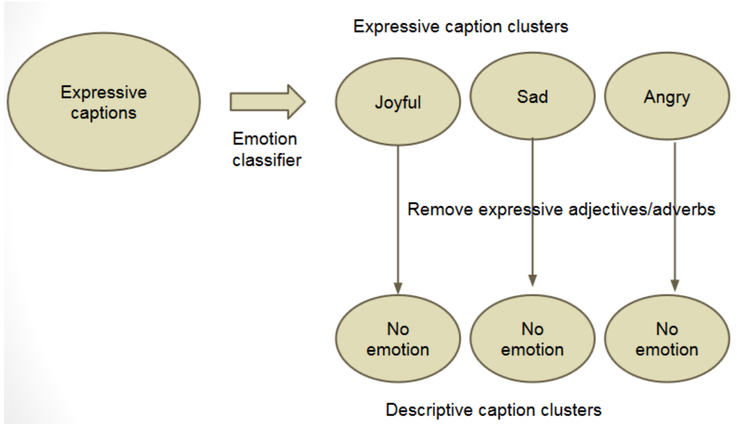
Someone, carrying a basket of readers' letters, comes out of someone's office with a happy smile.



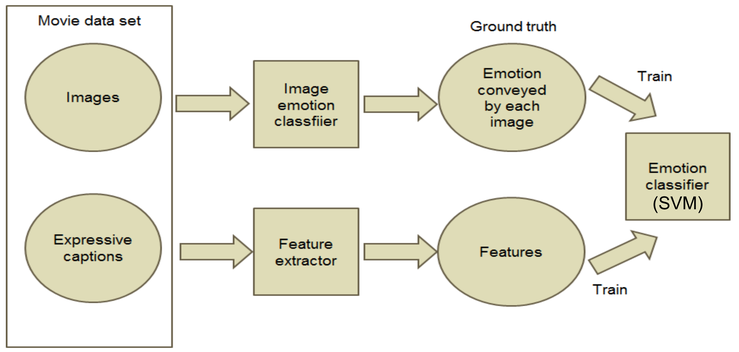
His eyes are moist with tears.

Approach:

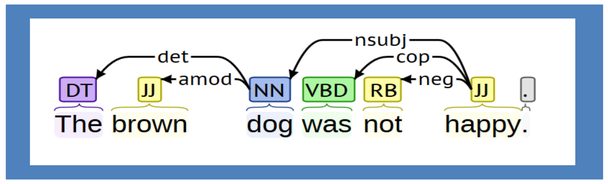
- Training
  - Divide expressive captions into clusters by emotion
  - Generate corresponding descriptive captions
  - Convert captions to vectors
  - Quadratic programming
- Testing
  - Classify emotion of descriptive caption
  - Compare caption emotion to image emotion
  - If different, add expressive modifier words (adjectives and adverbs)



Emotion Classifier:



- Features (based on method of [2]):
  - Emotion of individual word
  - Part of speech
  - Dependency tree features



Dependency tree features [3]

Emotion Classifier Results:

Sentence	Emotion
Someone is standing amidst the delirious crowd, smiling to himself, drinking in the adulation.	Joy
Someone's look of disbelief at his changing physique.	Surprise

Future Work:

- Convert captions into vectors
  - Skip-thought vectors [4]
- Train model to add expressive adjectives and adverbs to a given sentence
  - Quadratic programming

References:

[1] A. Rohrbach, M. Rohrbach, N. Tandon, and B. Schiele. A dataset for movie description. In CVPR, 2015.
[2] Ghazi, Diman, Diana Inkpen, and Stan Szpakowicz. "Prior and contextual emotion of words in sentential context." *Computer Speech & Language* 28.1 (2014): 76-92.
[3] Manning, Christopher D., Surdeanu, Mihai, Bauer, John, Finkel, Jenny, Bethard, Steven J., and McClosky, David. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*.
[4] Kiros, Ryan, et al. "Skip-Thought Vectors." *arXiv preprint arXiv:1506.06726* (2015).