USING EMOTIONAL CONTENT FROM AN IMAGE TO GENERATE A CAPTION OF THE IMAGE

By: Harman Singh & Kyle Wang

INTRO

- Using emotional content from images to generate captions.

- Aim of the work is to explore how emotional cues in images can be leveraged to generate more expressive and contextually relevant captions

- Emotional understanding in AI has the potential to enhance user experience, improve content recommendation systems, and enable more empathetic human-machine interactions

PREVIOUS WORKS

- Template
  - Can't generate variety

- Retrieval
  - Can't generate image-specific captions

DATASETS USED

- 2013 (FER-2013) dataset. Collected by Google search API.
  - It included images labeled with standard facial expression categories (happiness, sadness, fear, surprise, anger, disgust and neutral).
  - 35,887 examples (28,709 for training, 3589 for public and 3589 for private test)
  - Grayscale 48x48 images.
  - Postprocessed/Cleaned (removing some images that were fully black, with no details)
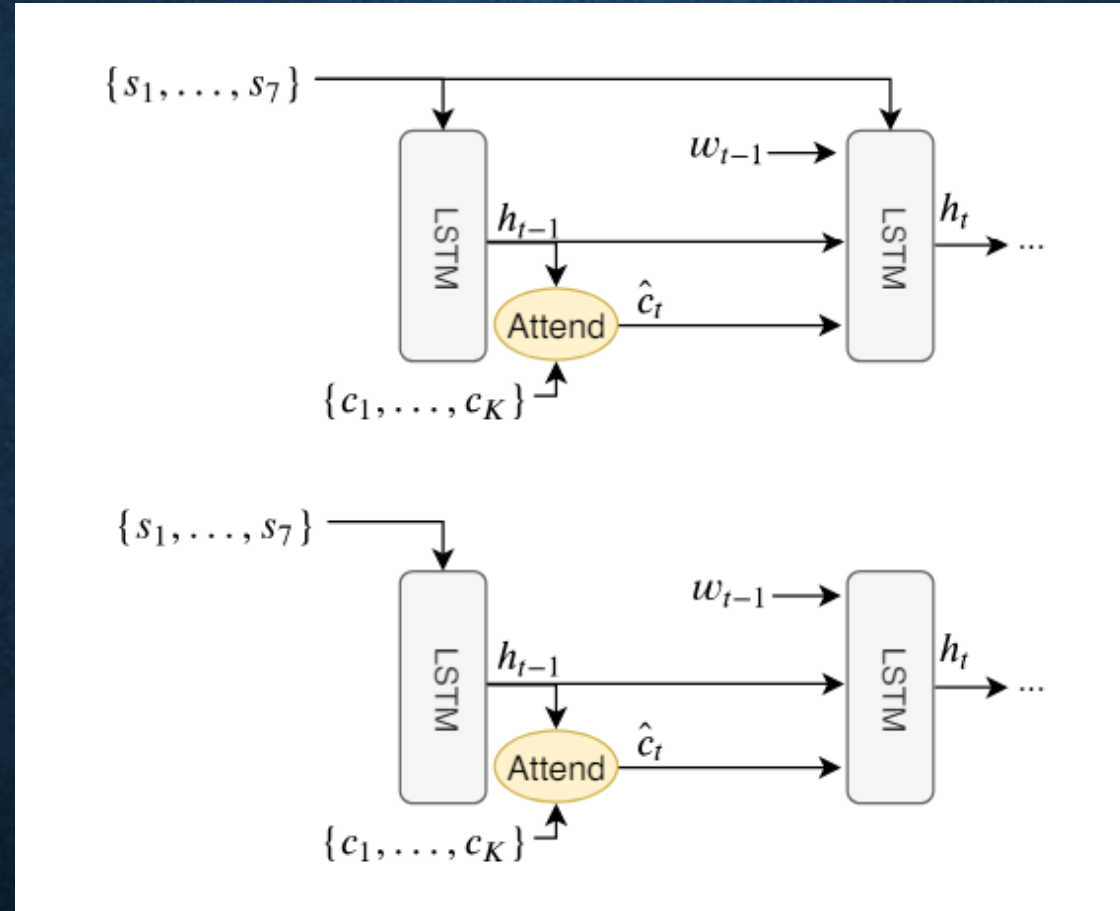
- FlickrFace11K dataset.
  - ~ 11,000 examples, 8696 for training, 2000 for validation and 1000 for testing.
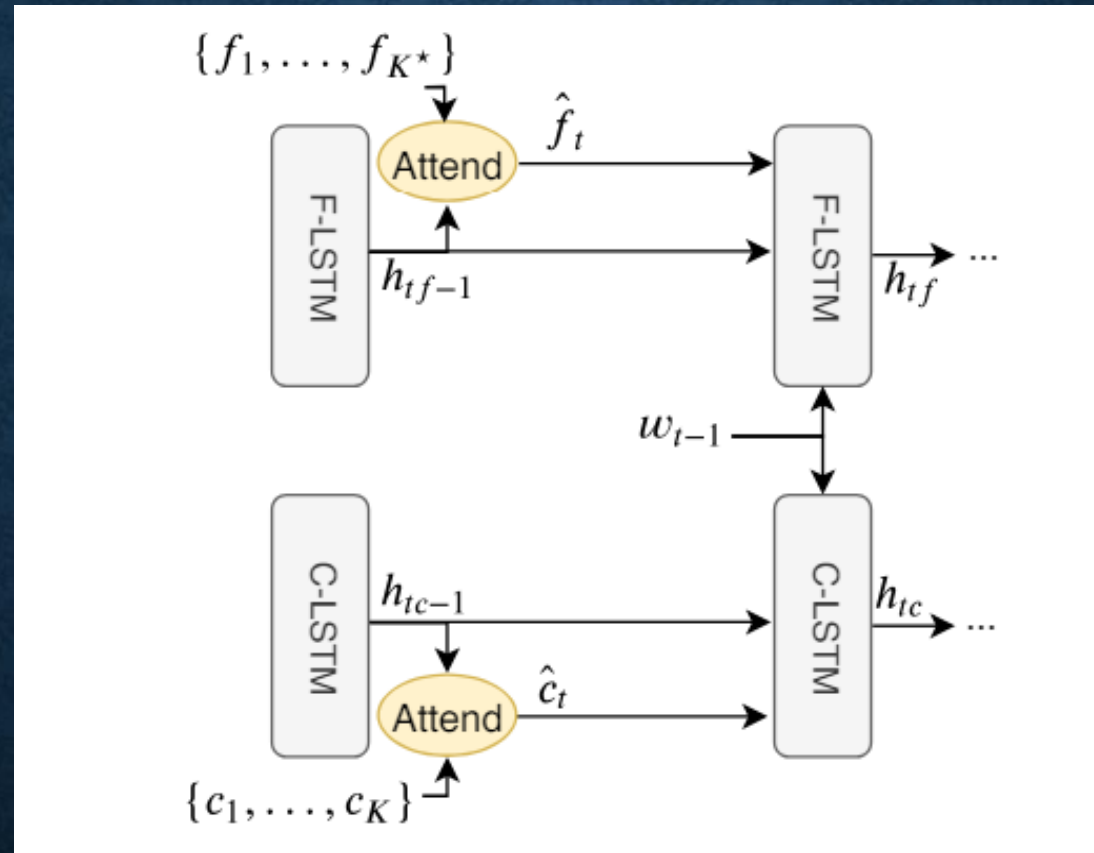 - Grayscale 48x48 images.

MAIN CONTRIBUTIONS

- Better model to caption emotional aspects of image

  o Model outputs more emotionally descriptive words

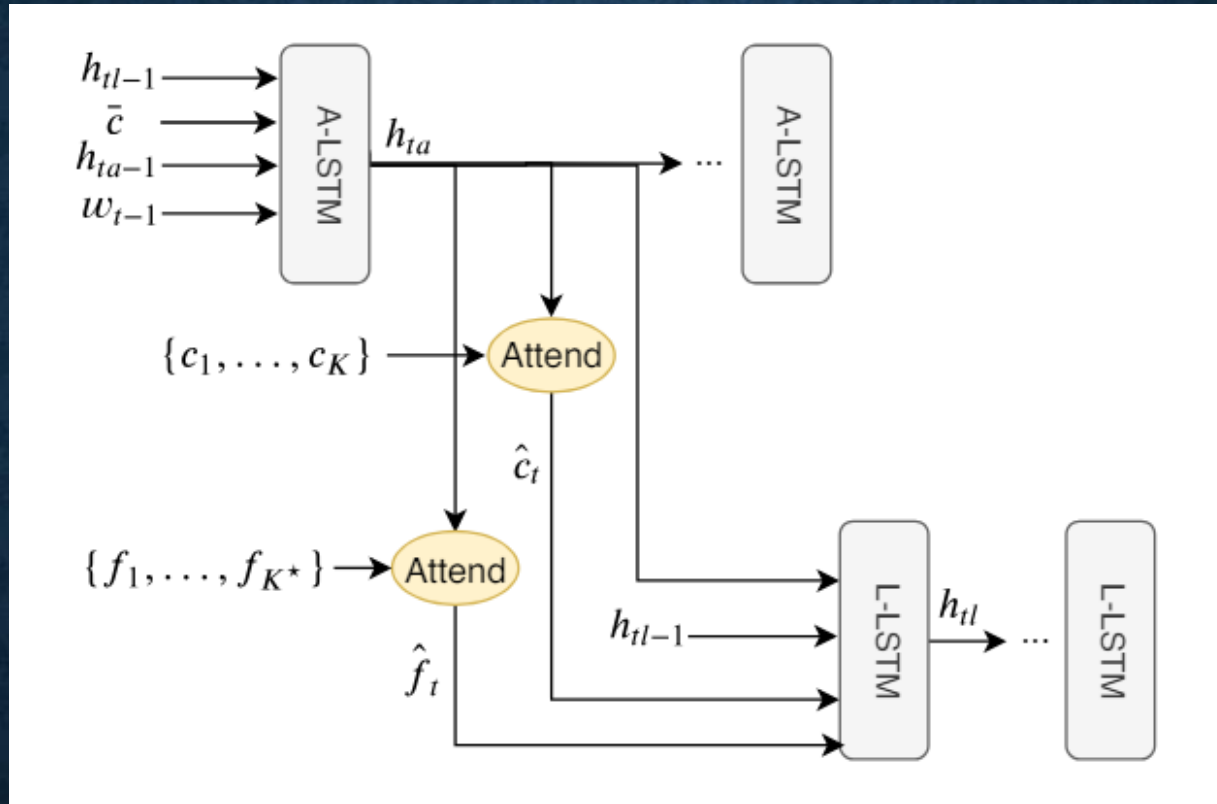- Model contributes to new techniques in AI

Difference between FACE-CAP-REPEAT and
FACE-CAP-MEMORY

DUAL-FACE-ATT models for facial features {f1 to fK} and visual content {c1 to cK}

JOINT-FACE-ATT model with two LSTMs that separately learn weights and generate captions

EVALUATION

- Although, it currently doesn't have much practical applications, its concepts are striking
  - o Being able to understand emotional context is a big feat
  - o Utilizes new self-attention models

FURTHER APPLICATION

- Visually impaired

- Automating tasks like social media, marketing, healthcare, and human-computer interaction.

- Larger scale impact: advancing the field of computer vision and natural language processing, particularly in the realm of understanding and generating human-like emotional responses.

- Qualifying emotional impact of image

LIVE DEMO

https://replicate.com/nohamoamary/image-captioning-with-visual-attention

DEMO GAME

young man jumps into the side

man in black cap and white shirt and white shirt and white shirt and white shirt and white shirt and white shirt and white shirt and white shirt and white shirt and

DEMO GAME

clown with stuffed toy

the child is on its mouth

QUESTIONS?