

UTILISATION OF ARTIFICIAL INTELLIGENCE FOR MULTIMODAL UNDERSTANDING OF MEMES

Aim: Training an Ai to be able to classify an image as either a meme or a non-meme

Introduction

In this day and age, memes are a common form of digital communication. However, despite their lighthearted intent, they have also been used to spread hateful messages across the internet. Inspired by the success of the Facebook Hateful Memes Challenge[1], this project delves into developing an Ai that can differentiate between memes and non-memes so as to pave the way to combatting hateful memes.

Related Works

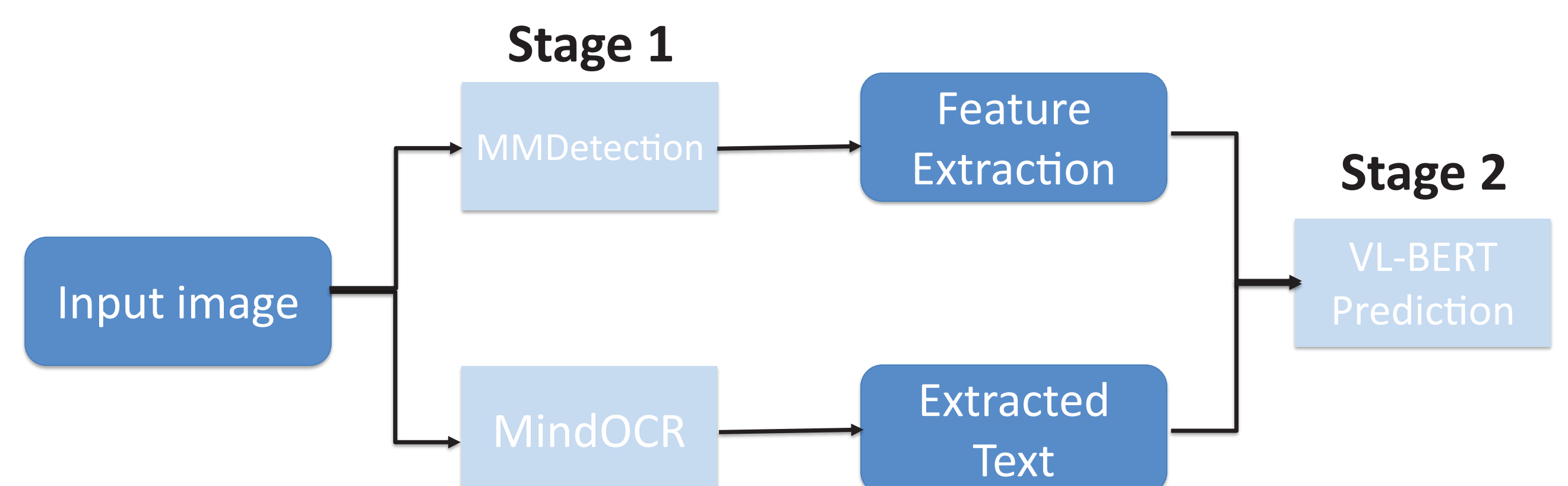
- Facebook Hateful Memes Challenge [1]
- Memotion 1.0 Task 8 [2]
- Detecting Sarcasm in Multimodal Social Platforms [3]
- Detecting Hate Speech in Multimodal Memes [4]

Results

Due to time constraints, VL-BERT was not fine-tuned on the final curated dataset and thus is unable to produce a prediction to classify an image as a meme or a non-meme. However, both MindOCR and MMDetection are able to perform their tasks with relative accuracy.



Method and Materials



- Model Pipeline

- **Stage 1:** Text and Feature Extraction
- **Stage 2:** Final Prediction

- Models Used

- **MindOCR** | Text Extraction using DB++ for detection and CRNN for recognition
- **MMDetection** | Feature Extraction for more accurate classification
- **VL-BERT** | Learning semantic association between image and text, crucial for differentiating memes from non-memes

- Datasets

- **MindOCR** : Total Text
- **MMDetection** : PascalVOC 2007 and 2012
- **VL-BERT** : Finetuned on VQA, RefCOCO+ followed by further finetuning from a curated dataset of memes and non-memes

Future Works

- Finish the finetuning of VL-BERT
- Further train MindOCR and MMDetection on more datasets to allow better detection on edge cases
- Allow for classification of memes into different categories such as Satire or Hateful using other models such as a Satire checker or Race and Gender classifier to assist in the analysis of memes
- Use Facebook Hateful Memes Challenge and Memotion Datasets to further finetune the model for more specific tasks

References

1. DrivenData. (2021). Hateful memes: Phase 2. Retrieved from <https://www.drivendata.org/competitions/64/hateful-memes/>
2. Sharma, C., Bhageria, D., Scott, W., PYKL, S., Das, A., Chakraborty, T., Pulabaigari, V., & Gamback, B. (2020). "SemEval-2020 Task 8: Memotion Analysis -- The Visuo-Lingual Metaphor!" arXiv preprint arXiv:2008.03781.
3. Schifanella, R., de Juan, P., Tetreault, J., & Cao, L. (2016). "Detecting Sarcasm in Multimodal Social Platforms." In Proceedings of the 24th ACM international conference on Multimedia (MM '16). ACM. DOI: 10.1145/2964284.2964321.
4. Kiela, D., Firooz, H., Mohan, A., Goswami, V., Singh, A., Ringshia, P., & Testuggine, D. (2021). "The Hateful Memes Challenge: Detecting Hate Speech in Multimodal Memes." arXiv preprint arXiv:2005.04790.
5. Title: I know most of us have strong feelings about cutlery but how about crockery. (n.d.). Reddit. Retrieved December 2023 from https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.reddit.com%2F%2Fautism%2Fcomments%2F137f1ykl%2Fi_know_most_of_us_have_strong_feelings_about%2F&psig=AOvVaw1xH1MMYQBF9grRXnb-V74&ust=1708950984215000&source=images&cd=vfe&opi=89978449&ved=OCBUQjhxqFwoTCiKjNTAxoQDFQA AAAAdAAAAABAE
6. Title: Hearing everyone get their coronavirus stimulus check and I'm just here like (n.d.). iFunny. Retrieved December 2023, from https://www.google.com/url?sa=i&url=https%3A%2F%2Fifunny.co%2Fpicture%2Fhearing-everyone-get-their-coronavirus-stimulus-check-and-i-m-js3rF9Ba7&psig=AOvVaw0csaE_KgL9468vKAJ8Y2TH&ust=1708951414279000&source=images&cd=vfe&opi=89978449&ved=OCBUQjhxqFwoTCiKjNTAxoQDFQA AAAAdAAAAABAE

Member:
Kaitlyn Janine Ang En, Raffles Girls' School
Mentors:
Kuek Yong Jie Adriel, DSO National Laboratories
Wong Minn Xuan, DSO National Laboratories