Fake News: Can we find any Consistency?

Introduction & Impact

The dissemination of fake news can exploit the sentiments of innocent people and societies, and serve as a catalyst for riots and violence. Social media has made it quite easy to spread fake news and manipulate information. Today, a piece of news with high-quality manipulations appears to be highly credible and accurate based on the factors like images resurfacing from past events with *mutation of visual evidence* and matter-of-fact details spread to support a fake text content, and high consistency between *Caption-text* and *image*. In such circumstances, it is only necessary to pay close attention to the details and entity relationships while fact-checking to identify fake news. Prior works [3, 2] focused on automatic misinformation detection and fact-checking by developing various benchmark datasets. The solution to this problem lies in our proposed 'multi-level context and relationship consistency check' framework evaluates the image-caption pair-based consistencies scores on various levels, taken from the news circulated through diverse mediums on social media platforms. As per the problem statement, we aim to propose this novel architecture for an investigative framework that performs deep analysis to find consistencies by applying logical and analytical reasoning, combining relationship understanding between entities at each stage.

Dataset Used

We make use of the NewsCLIPpings[1] dataset, which includes both authentic and fake images. It is based on the VisualNews corpus, which includes news articles from The Washington Post, The Guardian, USA Today, and the BBC. Different to previous misinformation datasets, in NewsCLIPpings both the images and captions are unmanipulated, but some of them are mismatched. A balanced dataset consists of 71,072 train, 7,024 validation, and 7,264 test cases. Each news in the dataset consists of a JSON file which includes the domain of news, image path, caption, URL of news, and news title.
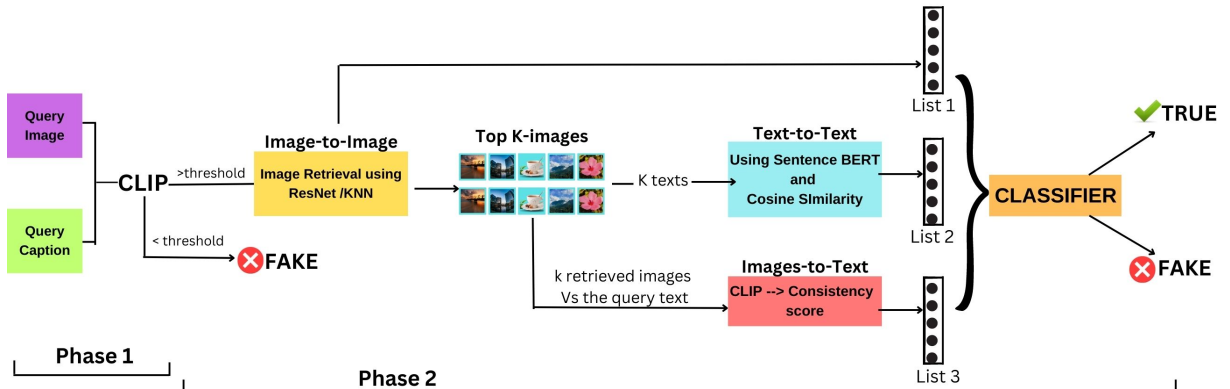

Figure 1: Proposed Two-tier architecture

Methodology

Our architecture suggests a consistency-check paradigm with two tiers. In first phase, the query image-caption pair's consistency scores are determined using CLIP. CLIP [1] is a large body of work on zero-shot transfer, natural language supervision, and multi-modal learning. The poor consistency scores represent the evident manipulations and re-purposed news images. News articles that have consistency scores above the cutoff are sent to the second-tier for in-depth study. On the secondary tier, similar images and their associated news captions are retrieved using reverse image search techniques. These retrieved news image captions are then compared with the query image caption in a cyclic way. The image-image similarity, image-text similarity and query text-to-text associated with retrieved image similarity scores are calculated.These calculated image-image, text-text, and image-text consistency scores are then fed into a classifier that forecasts the pristine or falsified label.

Preliminary Results

For every news article, we provide a consistency score ($0 \leq Cc \leq 1$) computed based on the above pipeline. These scores are used by classifier to predict labels, as depicted in Fig 1. In the first tier, the CLIP model after fine-tuning returns consistency scores of consistent images in the range 0.20-0.30 during the preliminary results. This is an ongoing project and we are currently working on the final results of proposed model.

References

[1] Grace Luo, Trevor Darrell, and Anna Rohrbach. Newsclippings: Automatic generation of out-of-context multimodal media. *arXiv preprint arXiv:2104.05893*, 2021.

[2] Shubha Mishra, Piyush Shukla, and Ratish Agarwal. Analyzing machine learning enabled fake news detection techniques for diversified datasets. *Wireless Communications and Mobile Computing*, 2022:1–18, 2022.

[3] Barry Menglong Yao, Aditya Shah, Lichao Sun, Jin-Hee Cho, and Lifu Huang. End-to-end multimodal fact-checking and explanation generation: A challenging dataset and models. *arXiv preprint arXiv:2205.12487*, 2022.

---

[1] https://openai.com/research/clip