Synthesis of Computer Vision and Website Design Principles for Misinformation-Screening Website

A Technical Report submitted to the Department of Computer Science

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Abigail Morgan

Spring, 2022

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Daniel Graham, Department of Computer Science

Synthesis of Computer Vision and Website Design Principles for Misinformation-Screening Website

CS 4991 Capstone Report, 2022

Abigail Morgan
Computer Science
The University of Virginia
School of Engineering and Applied Science
Charlottesville, Virginia USA
am5zm@virginia.edu

Abstract

The Internet allows more connection than ever before, but it also allows misinformation to spread at an unprecedented rate—via social media platforms, for example. Many social media platforms screen user-generated content for the purpose of flagging misinformation, but in some cases the fact-checking algorithms only focus on text-based content, not images or videos. Moreover, the public usually only interacts with algorithmic fact-checking in environments where they are one step removed from the fact-checker: many social media sites employ fact-checkers, but there is no way for users to submit posts to a fact-checker on their own. By combining skills learned from CS 3240: Advanced Software Design and CS 4501: Computer Vision, two upper-level computer science courses offered at the University of Virginia, a developer could create a website that screens uploaded images for misinformation. Advanced Software Design covers best practices in web development, facilitating the development of the website, and the image processing techniques covered in Computer Vision would allow the site to extract text from images, which could then be screened for misinformation.

1. Introduction

During the COVID-19 pandemic, the world faced two battles: the battle against the virus itself and the battle against misinformation. The nature of the connected modern world allowed information to spread more rapidly and to more people than ever before. In 2020, 230 million Americans (about 70% of the population) were regular users of social media platforms, which are a major avenue for the spread of misinformation [1]. About two-thirds of American adults get at least some news from social media, and about 40% of social media users say they expect the news they see on social media to be largely accurate [2].

Many social media sites use algorithmic fact-checking, in which user-generated content is screened and flagged if it meets certain criteria; for example, content might be flagged if it contains certain words or phrases associated with a topic that is prone to generating misinformation. Despite these efforts, however, misinformation continues to proliferate at an alarming rate. Allowing users direct access to algorithmic fact-checking, rather than only automatically screening content, would be an additional tool in the fight against misinformation. A website that screens user-uploaded content for misinformation using computer vision would accomplish this purpose.

Two computer science courses offered at UVA, CS 3240: Advanced Software Design and CS 4501: Computer Vision, provide the skills needed to create this website. Computer Vision teaches image processing techniques, some involving machine learning. Advanced Software Design covers best practices for agile software development; students work in a group to develop a website over the course

of the semester. Coupling these ideas would allow for the development of the proposed site and would contribute to the use of skills learned at UVA in solving real and important problems.

2. Related Works

In 2019, researchers at the International Conference on Intelligent User Interfaces presented Verifi2, a visual analytic system for investigating misinformation on social media [3]. This system allows users to explore news on both a source and document level and aids them in recognizing misinformation by highlighting "linguistic, network, and image features that can distinguish suspicious news accounts"[3]. While this paper focuses specifically on the ways in which Advanced Software Design and Computer Vision provide knowledge that is useful for constructing such a tool, the design of Verifi2 is an excellent example of what features such a tool would need in order to be useful and successful at aiding users in analyzing media they encounter. This research also provides an idea of the expertise and resources necessary to create such a system.

There are many browser extensions that analyze news content on either a source or document level [4]. The proposed website's functionality could be extended by creating a browser plugin that automatically screens images that the user views; but this functionality goes beyond material covered in Advanced Software Design. The use of a browser plugin would reduce the effort needed to evaluate content, since users would not have to manually submit the content to the site, while simultaneously increasing access to algorithmic fact-checking. However, many of these extensions do not analyze the images that circulate around news and social media content, which are themselves a source of misinformation [3].

SurfSafe is a browser extension that shows its users the source of images they view online so they can determine if it is being taken out of context [5]. When a user views an image, SurfSafe scans its database of images from trusted news sources to find the source of the image [5]. Using an image database in addition to computer vision techniques to recognize images would enhance the scope and capability of the proposed website. Examining the current landscape of algorithmic fact-checking would allow the developers of the proposed site to refine and enhance its capabilities.

3. Course Topics

The topics covered in both Advanced Software Design and Computer Vision provide important foundational knowledge that would be extremely useful in the creation of the proposed fact-checking website.

3.1 Advanced Software Development

In Advanced Software Design as offered in spring 2020, teams of students built a website over the course of the semester. As part of this project, students learned about the actual process of software development. Formalizing the process of creating software is extremely important, as it creates structure that makes development easier for independent developers and the teams that they work with. Agile development is a style of development that emphasizes iteration, allowing teams to check in frequently and as a result stay responsive to changing requirements.

Teams used Scrum, an agile development style. First, a list of product requirements is created and prioritized using client feedback [6]. Then, the team works in periods of 2-4 weeks called sprints, where requirements are first pulled off of the backlog during a planning meeting and then are worked on by the developers during the sprint [6]. Each day during the sprint, a short meeting called the daily scrum occurs, in which each member of the team shares what they are working on and any obstacles to their progress [6]. After the sprint ends, the team reviews the work they did and how it went, and then the process begins with a new sprint [6]. By pausing pure development between sprints and focusing on planning, teams stay responsive to changing requirements and are able to improve the way they work. This exposure to a development style often used in the software

development field prepares students for work in industry or on projects such as the proposed site.

Another way of adding structure to the development process is using source control management, another topic covered in Advanced Software Development. Source control management (SCM) is defined as "the practice of tracking and managing changes to code" [7], which aids development by keeping track of changes and allows them to be reversed if necessary. In Advanced Software Development, students worked in teams of five and used an SCM system called Git to manage the project codebase. Students were able to practice managing and combining their updates to the codebase, a skill necessary for development in industry that would also be important in the development of the proposed site. No matter the size of the team working on a project, SCM is a valuable tool in the pursuit of smooth and structured development.

Advanced Software Development also exposed students to web development languages and best practices. Students used Django, a Python web framework, alongside HTML and CSS to create their websites. This experience provides foundational knowledge necessary to build any website and thus helpful in developing the proposed site.

3.2 Computer Vision

Computer Vision as offered in spring 2021 did not specifically cover extracting text from images as the proposed site would, but rather introduced many different computer vision topics that provide more foundational knowledge. For example, students learned how to identify features of interest in images, such as lines and corners, which would be useful in recognizing letters since they are made of lines, corners, and curves. Students were also exposed to feature matching, which could also be useful in identifying letters; matching regions of images that contain text to images of letters can be used to extract text from images. Feature matching could also be useful in identifying images in order to provide their original context, like the SurfSafe browser extension discussed previously.

Students also used external computer vision and machine learning Python libraries to write their code, which provides useful experience in using existing tools to streamline development rather than building every piece of functionality from scratch.

3.3 Topics Not Covered

Some elements of the proposed website are beyond the scope of the topics covered in both classes. One of these areas is the actual fact-checking, which would involve determining if the submitted content contains misinformation or not. Despite this shortcoming, both courses expose students to topics essential for the basic development of the website.

4. Results

Both Advanced Software Development and Computer Vision provide students with important foundational knowledge that is important in industry and in the development of the proposed fact-checking website. The Department of Computer Science at UVA strives to "lead, explore and help define technologies, software and processes that will continue to change the world" [8], and preparing students for the future by teaching them relevant skills plays a large part in the success of that mission. The findings of this report affirm that CS course offerings at UVA provide useful skills that can help solve pressing and important problems like the rapid spread of misinformation.

5. Conclusion

The rampant spread of misinformation is a problem that must be confronted. Developing tools like the proposed fact-checking website provides additional ways to confront misinformation. The CS department at UVA offers classes, namely CS 3240: Advanced Software Development and CS 4501: Computer Vision, that prepare students to build important tools like the proposed website. Creating a curriculum that arms students with applicable skills that can be used in solving real-world problems benefits not only the students but also society in general.

6. Future Work

To expand this project, the proposed website could be actually created. UVA CS students who have taken both Advanced Software Development and Computer Vision would be good candidates as developers of this project since these two classes provide relevant experience. The functionality of the website could also be extended by developing a companion web browser extension.

References

[1] DataReportal. Digital 2020: The United States of America. (February 11, 2020). Retrieved February 20, 2022 from https://datareportal.com/reports/digital-2020-united-states-of-america

[2] Pew Research Center. News Use Across Social Media Platforms 2018. (September 10, 2018). Retrieved February 20, 2022 from https://www.pewresearch.org/journalism/2018/09/10/news-use-across-social-media-platforms-2018/

[3] Alireza Karduni, Isaac Cho, Ryan Wesslen, Sashank Santhanam, Svitlana Volkova, Dustin L. Arendt, Samira Shaikh, and Wenwen Dou. 2019. Vulnerable to Misinformation? Verifi!. arXiv:1807.09739v2. Retrieved from https://arxiv.org/abs/1807.09739v2

[4] Sebastian van der Lans. 12 Browser Extensions to Help You Detect and Avoid Fake News. (March 18, 2021). Retrieved February 19, 2022 from https://thetrustedweb.org/browser-extensions-to-detect-and-avoid-fake-news/

[5] Issie Lapowsky. This Browser Extension Is Like an AntiVirus for Fake Photos. (August 20, 2018). Retrieved February 19, 2022 from https://www.wired.com/story/surfsafe-browser-extension-save-you-from-fake-photos/

[6] Digité. What Is Scrum Methodology? & Scrum Project Management. Retrieved March 15, 2022 from https://www.digite.com/agile/scrum-methodology/

[7] AWS. What is Source Control?. Retrieved March 15, 2022 from https://aws.amazon.com/devops/source-control/#:~:text=Source%20control%20(or%20version%20control,merging%20contributions%20from%20multiple%20sources

[8] University of Virginia Department of Computer Science. Welcome to the Department of Computer Science. Retrieved March 16, 2022 from https://engineering.virginia.edu/departments/computer-science/about-computer-science/about-us#accordion1103060