Image Filtering and Universal Comment System

Akshay U
Dept. of Computer Science
College of Engineering Kidangoor, India

Arun Chandy
Dept. of Computer Science
College of Engineering Kidangoor, India

Rekha K S
Assistant Professor
Dept. of Computer Science
College of Engineering Kidangoor, India

Ananthu N
Dept. of Computer Science
College of Engineering Kidangoor, India

Nair Sanil Ramachandran
Dept. Of Computer Science
College of Engineering Kidangoor, India

***Abstract*: In today's world computers are connected by the vast network Internet. People surf across the internet to gain information and communicate with each other. The users log in to many of the famous websites online for this. These websites contain a vast area where various types of content are posted and are shared. Some of these contents are mature and can be accessed and used by adults only. It creates a problem for parents wishing to protect their children from such unsuitable content. To solve these problems, this paper presents a sensitive content filter system based on TensorFlow. Certain information even on trusted sites can be used to mislead the user viewing it. In such cases, a universal comment box can be added to such sites where the user can comment regarding the information. A comment interface in addition to it will also allow users to upvote & downvote such comment. The comment system will have spam protection to avoid spam comments and it will also have a voting system to prioritize good and useful comments. Sensitive content is filtered with up to 99% accuracy on all sites the user visits. Thus, our work helps to make the internet safer and more reliable.**

Keywords*: *Image filtering; Universal Comment System.

I. INTRODUCTION

The technological advancement has drastically reformed the aspects of human civilization in the past few decades. With the evolution of the internet and its applications, smartphones and social media are becoming more ubiquitous and are introduced at younger ages. Information is easily delivered all over the world through the internet, and regardless, it carries a negative or positive message. Information is accessible to anyone at any time. People use the internet to obtain information and to communicate with one another. The dangers for young teenagers include risks towards: personal safety, privacy concerns, access to disturbing and inappropriate material, social isolation, and an increase in mental health concerns such as depression, anxiety, and poor sleep. There is also a significant risk of children becoming targets of cyberbullying or sexting. These websites contain a vast area where various types of contents are posted and are shared. Some of these contents are mature & can be accessed and used by adults only. Such mature and adult contents can affect the mind of a minor viewing or accessing it.

Some of these risks are a result of limited adult supervision and parents who are not adequately versed in online safety procedures. In order to benefit from the positive aspects of technology, children, along with their parents, need to be aware of the possible risks of social media and how to navigate them. Teaching technology has become a necessary part of a child's education.

In the proposed system, To filter such contents, a browser extension can be implemented which will scan such contents and filter it. A browser extension can access the content of a webpage before it is shown to the user and modify it before rendering. By adding the browser extension, users can filter out sensitive images and can make comments on any webpage on the internet. Certain information even on trusted sites can be used to mislead the user viewing it. In such cases, a universal comment box can be added in such sites where the user can comment regarding the information. Being able to comment on any webpage may reduce the number of fake articles. Sensitive content is filtered and an accuracy of upto 99% is obtained on all sites the user visits. This helps to make the internet safer and more reliable.

II. LITERATURE REVIEW

The Internet and smartphones have become an integral part of our lives. The internet has become an enormous platform where people of various places, age groups can interact & communicate with each other. Users use this space to get contents and consume data in the form of information. The circulation and traffic on the internet increases with the number of users. Users access the internet on various devices. Though many children access the internet using a computer that is shared by a family, a recent study states that two-third of the minor users have access to the internet with their personal devices [1]. Like every other thing, the internet also has its good side and its flip side. Among the contents that are usually posted on the web, there are various contents that should be only exposed to the adult users. This is because such effects have a greater rate of social effects. Such contents if exposed to the minor set of users can affect the mental and physical health of the teenagers accessing it [2].

Usually, the mature contents that are available on the web include pornographic images that are available freely on the websites. However, not every mature content is a pornographic image. There are both safe and unsafe images. Some images include backgrounds or objects that resemble the human skin which are generally included in the safe category whereas some images include naked human body and other sexual contents which are considered or marked as

unsafe [3]. Therefore, a mechanism which can identify adult contents and can mark the image as safe or unsafe is required. The proposed model uses various machine learning algorithms that can categorize images as NSFW (Not Safe for Work). This image filtering model is implemented by developing a browser extension using TensorFlow. Browser extension extracts the contents on a website and compares them with the unsafe images that are stored and updated in the server's database & provides a solution as a response hence making the website child-friendly [4].

Many people surf on the internet rely mainly on the news portals which feature articles about various events happening around us. While there are many trustworthy and credible websites which provide genuine information, there exist some websites which publish articles that claim false & misleading information. Though some counter-news contents are posted on other sites, the misleading information that is shared by the former site can lead to the creation of propaganda among the public [5].

Such misleading and false information on the websites needs to be called out by the users accessing it by giving them an interface to communicate. This can be implemented to every possible website by creating a browser extension using the Reddit comment system mechanism which will allow users to comment & interact with the website. The users accessing it can upvote, downvote such comments. The voting system permits only registered users to upvote (give a positive +1 vote) or to downvote (give a negative -1 vote) on posts and comments. Two factors influence the post's ranking position there: 1) time and 2) voting score [6].In this way, this comment system can provide the users an interface on websites to call out such misleading information and share their opinions.

III. METHODOLOGY

A. Image Filtering

Image filtering refers to the process of scanning all the images in every web page a user visits and then classifying it as safe or not. All images are blurred by default to avoid revealing any image before scanning it and they are shown once the scanning is done and the image is deemed safe.

For classifying the images, a JavaScript library called nsfw js, which allows including custom models and returns a result that is more suitable for the current use case. The library uses tensorflow internally. Models are community made and have been proven accurate to up to 93%. There is already an extension based on nsfw js that adds this functionality; however, our implementation uses different approaches on scanning, identifying and storing the image results.

The extension has 3 parts – content scripts that are injected into web pages, background scripts that work in the browser background, and popup page that provides the UI.

When a user visits a web page, the content script and style are injected. The style will immediately take effect, blurring all the images. The script will wait for the DOM to be ready and will start working once it is ready. The purpose of the content script is to scan for images, assign a unique id to the image and then send the URL and id of the image to the background script. It will also start a listener that listens to changes in the DOM such as insertion of an image and changing the src of an image and then act accordingly. The content script and background script communicate with each other via the port message api [3].

Once an image URL is received at the background script, it first checks if the image exists in the scanned array, if it exists the result is instantly sent back. If it doesn't exist in the scanned array, it then checks if it exists in the waiting queue, if it exists there then the id is appended to that particular item of the queue so that once the scanning is done then the result will be applied to all the images with the same URL. If the image doesn't exist in the waiting queue, then the URL is searched in the local database (IndexedDB), if a result is found then it is used, otherwise the image is added to a queue to be processed.

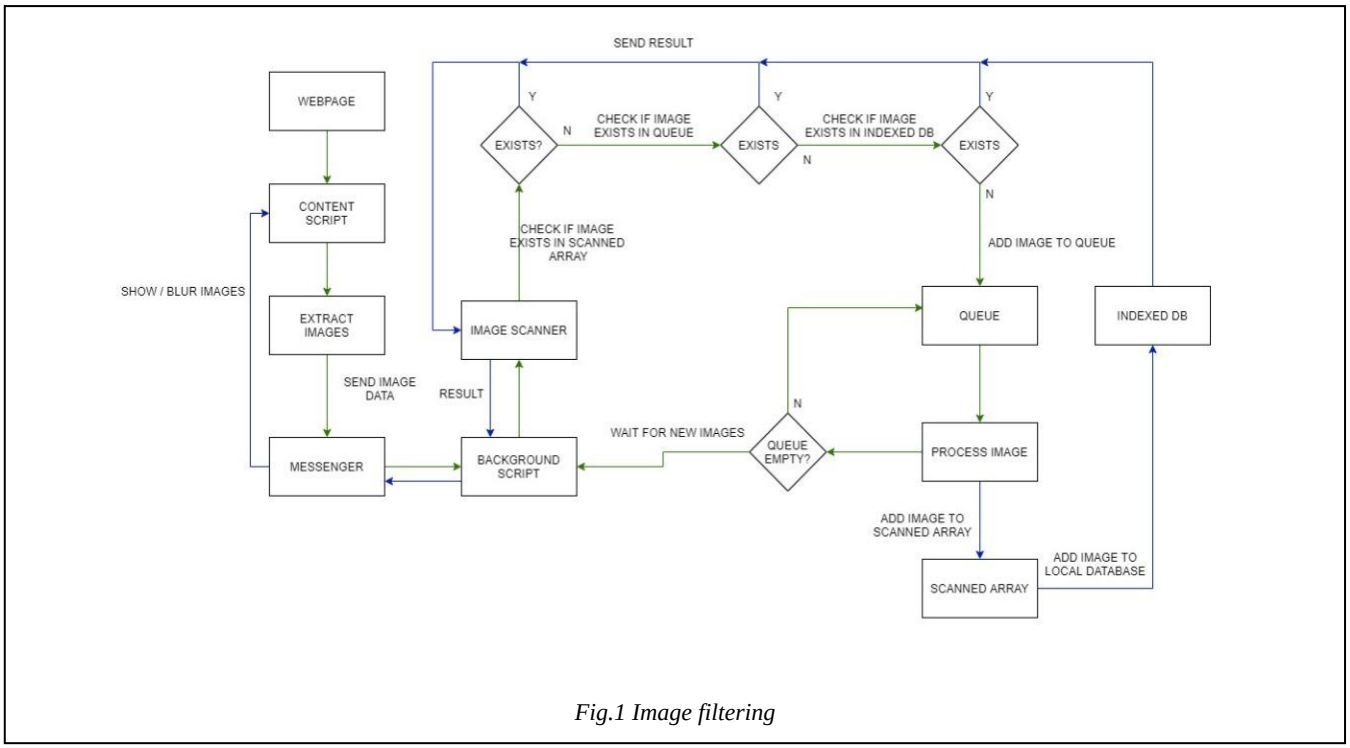

Fig.1 Image filtering

The queue is an array which stores the URL and ids of the images to be scanned. Image identification is done by removing the first item from the array and then scanning it. The result is appended to the scanned results and also stored to the local database. The process is recursively repeated until there are no more items in the queue.

When the content script receives a result from the background script, it will decide whether to reveal the image, moderately blur or keep it blurred based on the result.

Most false positive results are often in the moderately blurred area – this will allow the user to manually show or hide the image by right clicking and choosing blur or show the image, additionally when the user performs this action, the user's action is updated in the database. The popup UI allows the user to turn off the image filtering as well as provide some basic settings such as white-listing a website and backing up the local database.

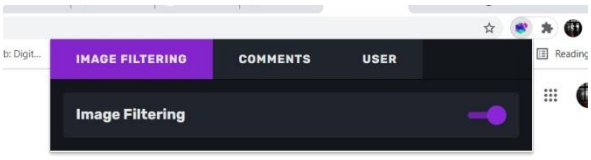

Fig : 3 Enabling Image Filtering

B. *Comment System*

The comment system provides a way to add your opinions about a page or article on any websites – including those that haven't implemented this functionality. Users can comment on a webpage by clicking the extension icon from the toolbar, which opens a popup page that provides the UI to comment, manage profile and other functionalities. A user needs to be registered and logged in to create and vote on comments, this is to reduce spam comments and vote manipulations.

When a user makes a comment, the comment, the webpage information and user information are sent to the server. The server will first ensure that the incoming comment is authenticated and then checks for spam, once done the comment is added to the database. The URL and web page information is stored along with comments. By default query parameters in the URL (text after '?', like ?search=abc) are ignored, however a user can opt to include the query string if they wish to do so – even in such cases some query strings like utm and ref are stripped.

Users can vote on a comment by clicking the Up arrow to upvote and down arrow to downvote. A comment score is attached to each comment based on the number of upvotes and downvotes it has received. Voting on your own comments is prohibited. Users can also report comments which will be handled by moderators. The server uses GraphQL api, and so other webpages and services can easily view comments using it.

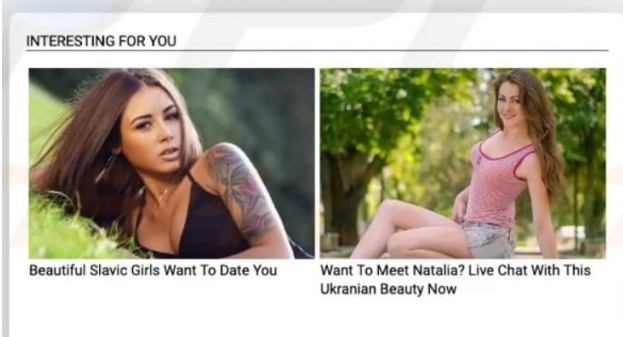

Fig. 2 Pop-up ad from a website.

IV. RESULTS AND DISCUSSIONS

Erotic and indecent content on the web is mainly concentrated on specialized sites that warn visitors about its "adult" content. Also, due to web spam and how the system works, this type of content can be shown to users even when they do not intentionally look for it. Even the most innocent websites may contain a porn banner that has ended up there by an advertising network

Fig. 2 is a pop-up ad from a website, exposure to such obscenity, violence, unmoderated chat sites, etc. can have fathomless effects both in the personal as well as professional life of the minors and has proved to be catastrophic in the past. The benefits of our extension are two-fold: It can actively filter and block indecent content, and it also provides a comment box with which a user can add opinions about a webpage or any article on websites. In addition to the comment interface, will also allow users to upvote and downvote comments.

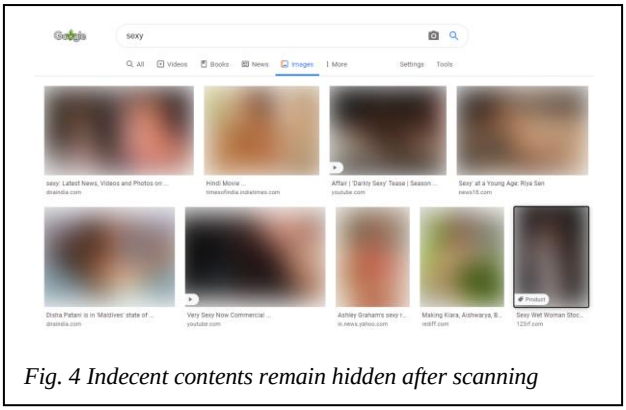

Fig. 4 Indecent contents remain hidden after scanning

The browser extension provides a safer browsing experience by filtering the sensitive images present on the webpage. After enabling the image filtering option as shown in fig 3, it blocks NSFW images from the web pages. Works without any backend servers by running everything in the user browser itself. We used a JavaScript library for classifying the images, the library uses TensorFlow internally.

For the implementation we use different approaches to scanning, identifying, and storing the image results. When the user visits a web page, all images are blurred by default. During scanning, if any offensive or adult images are found such contents get blurred. Fig 4 shows the final result of a scanned webpage, identified indecent content is filtered with

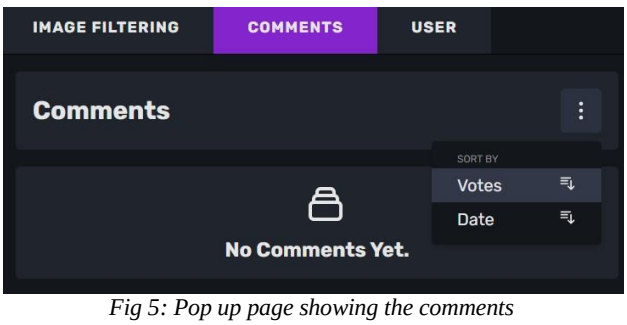

Fig 5: Pop up page showing the comments

up to 94% accuracy on all sites the user visits. The creation of a local database allows storing the results of scanned images and thereby improving the efficiency.

Ensures data privacy, we did NOT collect/send any user data. All the operations on the images are done locally on the browser. No user data is being sent to a server for processing.

Registered users can comment on a webpage and they can report spam or abusive contents. Reported comments will be evaluated by the moderators. The server uses GraphQL api, and so other webpages and services can easily view comments using it.

V. CONCLUSION AND FUTURE SCOPE

In this paper, the idea of a TensorFlow-based adult content detection system is proposed to automate the classification of adult content on the internet. Different sensibility issues are considered, including identification of the adult content. The proposed mechanism offers an adult content detection system that filters the mature content and makes it inaccessible to minors. Indeed, the proposed mechanism successfully implements a universal comment box on various sites which allows the user to comment his opinion about the article. This image filtering technique can systematically retrieve all the web content and analyze this content alongside various data models stored in the server's database to identify and block adult contents. It can also be used at home, in offices and schools, and in other public sectors to intelligently investigate a website. Furthermore, the comment system provides an user interface where the people visiting a website can vote the comments accordingly. In future work, the detection method for adult content will be further improved to be used by the users who are visually impaired. Such browser extensions can contribute to making the internet a safer place.

REFERENCES

[1] Dyer, Tobbi. "The effects of social media on children." Dalhousie Journal of Interdisciplinary Management 14 (2018).
[2] Ali, Farman, et al. "A fuzzy ontology and SVM–based Web content classification system." *IEEE Access* 5 (2017): 25781-25797.
[3] Devi, S. Anoosha, and V. Arvind. "Detection of Illegitimate Divulgation of Obscene Contents Using TensorFlow." International Journal of Applied Engineering Research 13.20 (2018): 14595-14599.
[4] Khan, Rehan Ullah, and Ali Alkhalifah. "Media Content Access: Image-based Filtering." International Journal of Advanced Computer Science and Applications 9.3 (2018): 415-419.
[5] Hopp, Toby, Patrick Ferrucci, and Chris J. Vargo. "Why Do People Share Ideologically Extreme, False, and Misleading Content on Social Media? A Self-Report and Trace Data–Based Analysis of Countermedia Content Dissemination on Facebook and Twitter."
[6] Medvedev, Alexey N., Renaud Lambiotte, and Jean-Charles Delvenne. "The anatomy of Reddit: An overview of academic research." *arXiv preprint arXiv:1810.10881* (2018).