Instagram Image Filtration with Computer Vision

Tanjimul Ahad Asif* and Baidya Nath Saha
Department of Mathematical and Physical Sciences, Concordia University of Edmonton, Alberta, T5B 4E4, Canada

* Corresponding author

doi:https://doi.org/10.21467/proceedings.115.22

ABSTRACT

Instagram is one of the famous and fast-growing media sharing platforms. Instagram allows users to share photos and videos with followers. There are plenty of ways to search for images on Instagram, but one of the most familiar ways is 'hashtag.' Hashtag search enables the users to find the precise search result on Instagram. However, there are no rules for using the hashtag; that is why it often does not match the uploaded image, and for this reason, Users are unable to find the relevant search results. This research aims to filter any human face images on search results based on hashtags on Instagram. Our study extends the author's [2] work by implementing image processing techniques that detect human faces and separate the identified images on search results based on hashtags using the face detection technique.

Keywords: Face detection; Image processing; Selfie; Instagram; hashtag; Haar Cascade classifier; HOG descriptor.

I. Introduction

There have been many studies in recent years that address human face detection with the Haar Cascade classifier. Adri Priadana and Muhammad Habibi study the correlation between subject matter and photo content on social media [2]. They perform web data extraction using beautiful soup to download Instagram images and detect selfies with the Haar Cascade Classifier. Besides, they used the OpenCV library to perform the Haar cascade classifier, which gives high accuracy and superior performance with a simple background image. However, filtering selfie face images on Instagram does not entirely provide the proper hashtag-based search result as images contain human figures other than just the selfie pictures. The original paper aims to filter selfie face images on search results based on hashtags on Instagram [2]. However, our proposed study extends the original paper's work to improve the results by implementing image processing techniques while detecting a human object and separate images on search results based on hashtags using the face detection and pedestrian detection technique. We use the Haar Cascade classifier and HOG descriptor, which is 64.6% accurate to detect the human figure in a picture.

II. Literature Review

The increasing use of technology has made social media an integral part of our everyday life in recent years. Social media are designed to share content instantly, efficiently, and in real-time [3]. People tend to enjoy connecting more with peoples from all around the world and share moments. As well as it makes internet users spend more time on social media than other platforms. Instagram is one of the popular photos and video sharing platform with its large number of users. Based on Tech-crunch data, Instagram has 1 billion monthly active user record as of June 2018, which has jumped from 800 million active users in September 2017 [1]. Instagram encourages users to share pictures and videos with followers. More than a billion pictures are posted to Instagram every day, which is very plentiful. The enormous amount of pictures and videos enables Instagram to be a reference for an insight into a tourist spot, location in a city or any favorite destination for people to look for [2]. Instagram permits users to add a caption to the uploaded pictures. Users mostly post images with a symbol called a hashtag as a caption. Hashtags

Proceedings DOI: 10.21467/proceedings.115; Series: AIJR Proceedings; ISSN: 2582-3922; ISBN: 978-81-947843-5-7

can help better accessibility to the content, increase interaction, draw followers, improve the brand and community's image to reach a target audience [4]. On the other hand, hashtag searches are the most popular and efficient for finding specific Instagram content [3]. However, there are no guidelines for using hashtags to the uploaded images. As a result, the hashtag provided is often not relevant to the images uploaded. It makes the key topics and the relation between the images linked to the related hashtag details are low [2]. This research attempts to use the face detection technique to separate the human face's pictures on hashtag-based search results on Instagram. The model uses a web data extraction technique using the open-source library Instagram scrapper to download Instagram photos using the hashtag and face detection techniques applying the Haar Cascade method to detect the human face. We use the Haar Cascade classifier to detect a human face because of its simplicity, efficiency, reliability, speed and real-time approach to the application [2].

III. Methodology

The suggested method consists of three key phases: Web Data Extraction, Haar Cascade Classifier and Human Face Detection. Figure 1 shows flowchart for human face detection from Instagram images.

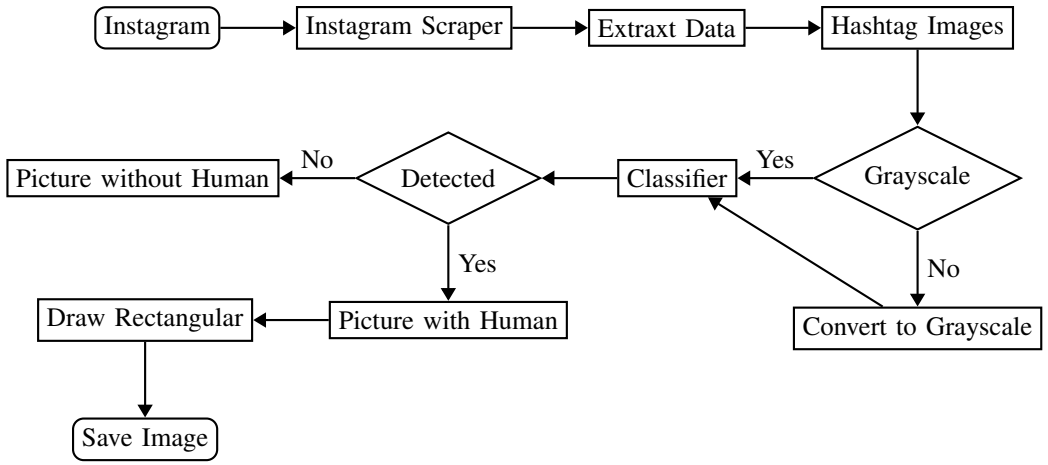

Fig. 1: Proposed methodology for human face detection from Instagram images.

A. Web Data Extraction

The Instagram images consist of three fundamental parts, mainly uploaded image, a caption, and hashtag. Hashtags addresses with symbols #, which are used to index the subject on Instagram. This function enables users to search for the stuff that they are interested in. The following Figure demonstrates the basic anatomy of a search using a hashtag. The data collection process uses the open-source python library Instagram-scrapper to extract image data from Instagram. Instagram scrapper downloads images with the given hashtag and stores it into a directory with the hashtag's name.

B. Cascade Classifier

In this project, we implement Haar Cascade Classifier to detect different human facial parts with OpenCV. We also implemented HOG descriptor to detect human body. The Haar features are the central part and use to identify elements in the image. Haar feature uses the rectangle features to detect human face rapidly. The features are calculated by deducting the number of pixels in the black square from the number of pixels under the white square as illustrated in Figure 2. The Haar cascade classifier uses the processed image and detects the face in that image.

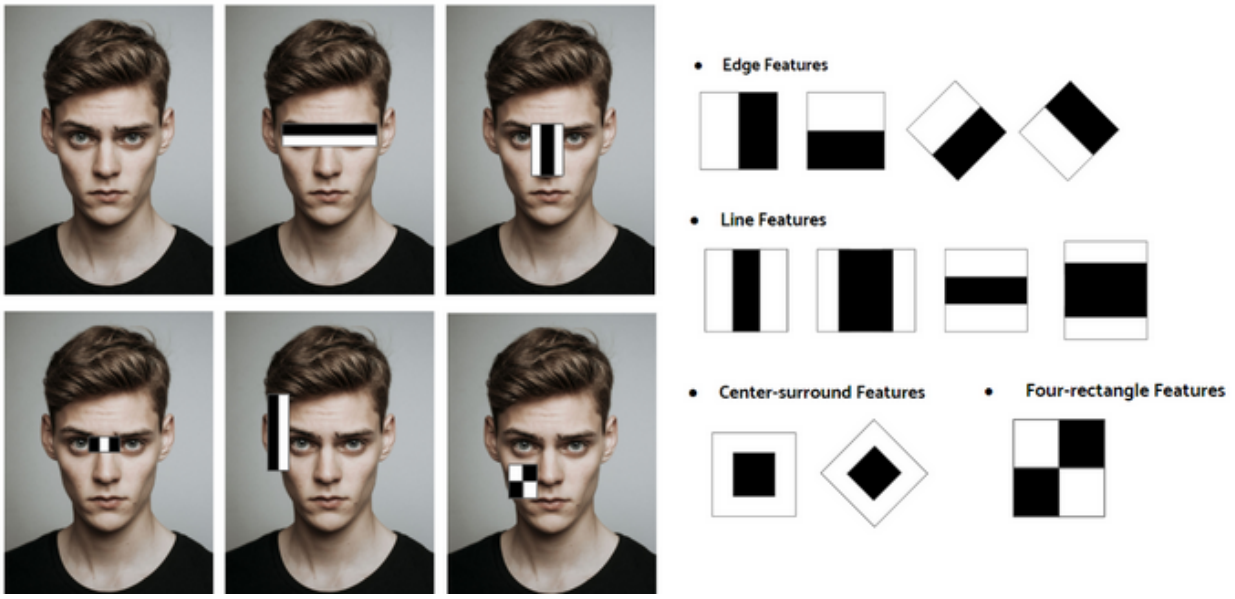

Fig. 2: Haar like features

C. Face Detection

Lastly, the Haar cascade classifier method is applied to every single processed image. If it detects any human face on the image, the application draws a rectangle around the face. Also, the image will automatically be filtered and separate the photo into a different directory. Figure 3 illustrates face detection using Haar cascade classifier.

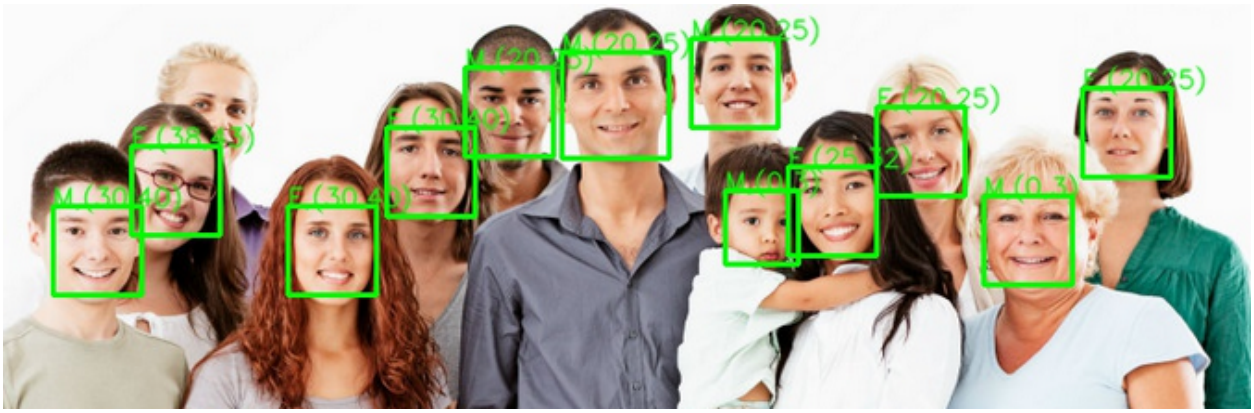

Fig. 3: Object detection with classifiers

IV. Experimental Results

The application is developed with Python programming language, incorporated with the OpenCV on a laptop computer with Intel 8-Core Processor @ 2.40 GHz CPU and the RAM is 16GB. It needs minutes to process 100 Instagram images and detect human figures. The original paper only applied default face-cascade to detect the human face. So, the application does not work for any other human body parts. On the other hand, in this project, we implement all the cascades for different human body part detection that produce an extraordinary result as demonstrated in Figure 4. The experiment was carried out by determining several hashtags as the basis for image search on Instagram. Table I shows the accuracy of different datasets.

Proceedings DOI: 10.21467/proceedings.115
ISBN: 978-81-947843-5-7

Series: AIJR Proceedings
ISSN: 2582-3922

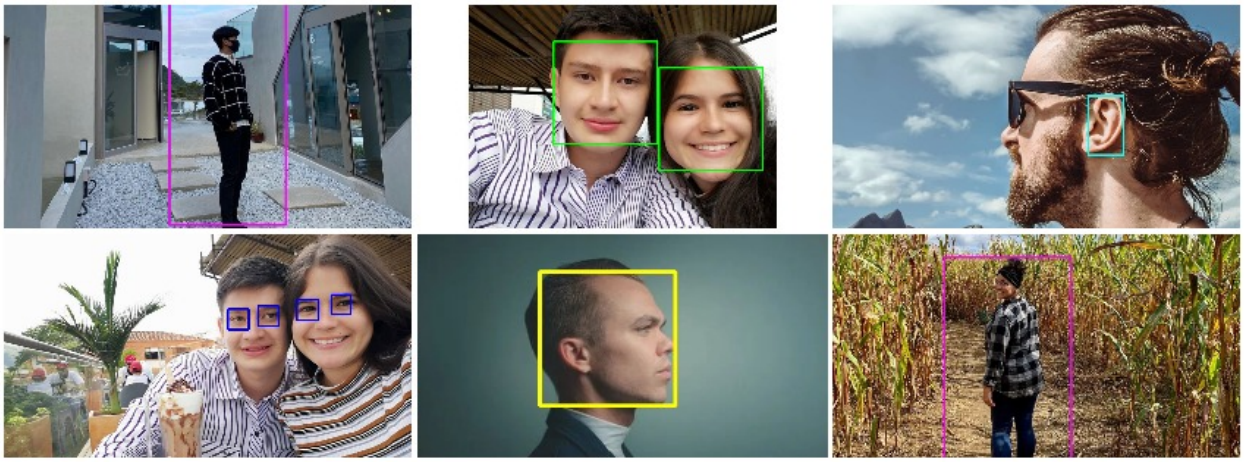

Fig. 4: Detecting different human body parts

Dataset	Positive Image	Negative Image	Accuracy
Selfie Picture	200	200	93%
Group Picture	200	200	72%
Upper Body	100	100	56%
Lower Body	100	100	48%
Profile Face Photo	100	100	86%

TABLE I: Accuracy of Haar like features

V. Conclusion and Future Work

This study analyzed Instagram images and detected human pictures using hashtags to provide precise search results. This work's implementation enhances the process of identifying human pictures using the HOG descriptor and Haar cascade classifier.However, the Haar cascade accuracy value is 64.6%, so it may face difficulty detecting the face in some cases. We plan to extend this work by training the classifier to improve our future work's accuracy value.

VI. References

1. "Techcrunch," 2019. [Online]. Available: https://techcrunch.com/2018/06/20/instagram-1- billion-users/

2. A. Priadana and M. Habibi, "Face Detection using Haar Cascades," in IEEE, Yogyakarta, Indonesia, Indonesia, 2019

3. L. M. S. K. Yuheng Hu, "What we Instagram: A first analysis of Instagram photo content and user types," in ICWSM, Ann Arbor, United States, 2014.

4. https://link.springer.com/article/10.1007/s11747-019-00695-1

5. OpenCV: https://docs.opencv.org/3.4/db/d28/tutorial_cascade_classifier.html