Containing the spread of Fake Images using Computer Vision and Image Processing

Piyush Aggarwal

The Indian Express, Noida, India
Me.PiyushAggarwal@gmail.com

Abstract. We are living in a world of easy media manipulation [1]. The online conversational spaces worldwide are swamped with disinformation and biases from every nook and corner of the society. On top of that, the role played by the social networks in the circulation and consumption of news has profound implications in terms of our media consumption and, in turn, the way public opinion is shaped by encouraging increased political polarization. Lack of neutrality of platforms with their hard to regulate algorithms, which are constantly manipulating our feeds, is the biggest obstacle every newsroom across the globe is facing today. Exponential developments in technology have made the creation and spread of false content lightning fast, which calls for immediate action to combat this menace and the development of necessary tools and apps to contain its spread. In this paper, we are proposing an image-matching algorithm, to find near duplicates of a given image from a defined dataset of images, which are saved in a database along with their characterizations and have already been fact-checked by the trusted fact checkers.

Keywords: Fake News, Fake Images, Fact Checking, Computer Vision, Image Processing, Similar Images, Near Duplicates

1 Introduction

Most misinformation spreads in the form of images and videos as the visual representation has more appeal and makes a stronger and lasting impression. At any given time, a majority of such information in the ecosystem are things that have already been debunked, however, they are in recirculation. Moreover, images are increasingly becoming a medium by which people communicate, as advancements in technology have enabled an exponential surge in taking images. Recently more than 50% percent of Twitter impressions are the posts with images, video or other media [2]. We need to develop smart and intelligent infrastructure to tackle this and to prevent our political discourse from getting fractured. As there is a glut of tools already available, therefore, we need to strengthen the existing infrastructure and distribution network. Not all of this is always possible with only web and social data, as we need to discov-

er other sources of data and information and tap into the psychological and contextual metrics, which are sometimes hidden in plain sight.

Here, we are proposing an algorithm to combat and stop the further dissemination of the fake news spread by fake images, which are already debunked by trusted fact checkers. Our proposed work mainly focuses on turning fact-checked images into a searchable database, where any user can input an image, which then can be compared with the existing images in the database using image matching and image processing techniques. Our aim is to be able to detect the closest match/ near duplicate of the input image, even, if it has undergone several transformations.

2 Existing Work

There exists many techniques and algorithms especially developed for finding similar images or near duplicates in a given dataset of images or for a given test image. For instance, RIME [6], the first near duplicate image search algorithm used wavelet coefficients of a given image to find its nearest match in an existing dataset. Connor et al. [7] used different distance metrics over various image characterizations to create clusters of near duplicate images. Karami et al. [8] compared SIFT, SURF and ORB techniques for finding the closest match of an image, undergone transformation and deformation. Morra et al. [9] have done an exhaustive work on benchmarking all such unsupervised methods for finding duplicate images.

We observed, though, there are many ways for finding near duplicates, however, the technique used largely depends on the application under consideration. For example, it could be for finding the best photo among several duplicates, for memory and search query optimization, to cluster similar images in a dataset, etc. [7, 9]. Moreover, datasets used in different methods are also different from one another and specialized for a given use-case. After considering such observations, we realized that developing a generic algorithm for finding near duplicates is a challenging task as finding common features, which can work on any type of image, can be very arduous.

3 Proposed Algorithm

We put together our dataset from publicly available images, fact checked by trusted fact checkers in India. While analyzing the dataset, we found the human face was the most common entity in 98% of the images. As for the forged region, it majorly involved the activity the person is doing, the object the person is holding or something with the background, etc. Therefore, we designed our algorithm using face as a primary deterrent to look for a given image in our dataset.

We start with the Perceptual Hashing technique [5] to choose the closest set of matched images from our existing database for the given input image. This preprocessing step not only provides adequate results in terms of images being present in

98% of the cases, however, also reduces the time complexity by avoiding unwanted computations on the remaining images in the database.

After narrowing down our database to the selected few images, we propose a face detection and matching technique between the selected images and the given input image. Face detection and matching technique will further reduce the selected images for further processing based on the number of detected and matched faces between the input image and the selected images. For cases where the input image has no face, we propose to use a label matching technique. Similarly, the label matching technique will reduce the number of images for final processing. Finally, we propose to apply the SIFT [3] and structural similarity (SSIM) index [4] on the remaining images to find the best match for the given input image.

4 Results

In order to test our proposed algorithm on the publicly available datasets, we could not find any relevant datasets, which we can refer. We explored popular datasets [9], however, they were not suitable for our use-case as they either had natural scenes (INRIA Holidays dataset) or had indoor/outdoor images from commercial and residential buildings (CLAIMS dataset) or had various photography images (MFND). The California-ND dataset was the closest dataset we could use, however, the faces were blurred so we could not use the same.

We tested our algorithm on 710 images and so far, have achieved an accuracy of 96%, which we calculated as mentioned below:

$$\text{Accuracy} = (\text{True Positives}*100)/\text{Total Images}$$

5 Conclusion and Future Work

Our aim is to be able to detect the closest match of the input image, even, if it has undergone several transformations. We have tested and verified the proposed algorithm under different types of transformations e.g. cropped, flipped, blurred, enhanced, and gray scaled, etc. and their multiple combinations applied to an input image along with reduced time complexity. We plan to work towards the detection of deep fakes in the foreseeable future by developing an intelligent model to detect if an image has been forged, manipulated and be able to detect the local forged, manipulated region using digital image forensic methodologies and deep learning.

References

1. Andrew J. O'Keefe: Welcome to the New Era of Easy Media Manipulation. https://singularityhub.com/2016/11/13/welcome-to-the-new-era-of-easy-media-manipulation/ (August 2019).

2. Mary Meeker: Internet Trends Report 2019. https://www.bondcap.com/report/itr19/ (August 2019).
3. Lowe, David G.: Object recognition from local scale-invariant features. *Proceedings of the International Conference on Computer Vision*. 2. pp. 1150–1157 (1999).
4. Wang, Zhou., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*. 13 (4): 600–612 (2004).
5. Zauner, Christoph: Implementation and Benchmarking of Perceptual Image Hash Functions. *Master's thesis, Upper Austria University of Applied Sceinces, Hagenberg Campus* (2010).
6. Chang, E.Y., Wang, J.Z., Li, C.: RIME: A replicated image detector for the world-wide-web. *Proceedings of SPIE Symposium of Voice, Video, and Data Communications* (1998).
7. Connor, R., MacKenzie-Leigh, S., Cardillo, F. A., & Moss, R.: Identification of MIR-Flickr near-duplicate images: a benchmark collection for near-duplicate detection. In *10th International Conference on Computer Vision Theory and Applications VISAPP* (2015).
8. Karami, E., Prasad, S., & Shehata, M.: Image matching using SIFT, SURF, BRIEF and ORB: performance comparison for distorted images. *Newfoundland Electrical and Computer Engineering Conference* (2015).
9. Morra, L., & Lamberti, F.: Benchmarking unsupervised near-duplicate image detection. In *Expert Systems with Applications* (2019).