# Containing the Spread of Fake Images using Computer Vision and Image Processing

Piyush Aggarwal The Indian Express, India

@MePiyush



## Fake Image - What

- Distorts truth
- Shows things never happened
- People say or do stuff they never said or did



#### Fake Image - Sample Image (Fake)





## Fake Image - Sample Image (Real)





## Fake Image - Sample Image (Fake)





## Fake Image - Sample Image (Real)





#### Fake Image - Sample Images (Fake)





## Fake Image - Sample Images (Real)





## Fake Image - Why

- Seeing is believing
- Pic or it didn't happen
- Photography furnishes evidence\*
- Conformational bias



## Fake Image - How

- Splicing
- Copy Paste/Move
- Inpaint/Interpolation
- Resampling
- Cloning
- Deep fakes
- Misappropriation (real image)



## Fake Image - Virality

- Engaging
  - Most mis-, dis-, and mal-information is in the form of images and videos
- Recirculation
  - Majority of content are things that have already been debunked
- Under resourced fact-checkers



## **Our Work - What**

 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



## **Our Work - Why**

- Information chaos is a design feature of platforms
- There are many tools already available
- Repetitive work by fact-checkers
- Need to strengthen and improve existing infrastructure and distribution
- Improves productivity of fact-checkers



## **Our Work - How**

- Perceptual Hashing
  - Narrows down the search and reduces time complexity
- Face/label detection and matching
  - Further reduces the matches
- SIFT & SSIM for best match on remaining images



## Similar Work

- Many techniques for finding similar images
- Need framework to evaluate published methods
- Different applications and context
- Various specialized datasets



## Results

- Tested on 710 images
  - Included cropped, flipped, blurred, enhanced, and gray scaled, etc.
- 96% True Positives
  - Accuracy = (True Positives\*100)/Total Images



















#### **Results - Failure**





#### **Results - Failure**





## Future

- Will be releasing an API
  - API can be plugged into platforms, websites, apps or turned into bots, a browser extension, etc. to check whether an image or meme has already been fact-checked automatically
- Model to detect if image has been forged or not and to detect the local forged region/s



## **Thank You!**

**Piyush Aggarwal** The Indian Express, India

@MePiyush

