New Perspectives on Digital Media Integrity

Walter J. Scheirer
Department of Computer Science and Engineering
University of Notre Dame
“The battle for the mind of North America will be fought in the video arena”

Brian Oblivion, Videodrome
Videodrome (1983)
Deep Fakes
ISIS Execution Videos (Fake)

The Intersection of Politics and Entertainment: Internet Memes
What is a Meme?

Memes, Visual Style, and Semantic Analysis

A naive approach:

Original

Dense grid of HOG descriptors

K-Nearest Neighbors
Related Work

Memes as Cultural Artifacts


Visual Style in Computer Vision


Empirical Effects of Manipulated Media

Memes and Political Interference
Brazil

Bolsonaro’s actual fraudulent employee: his brain

The girl you like

Her father

You better get used to him

Her ex

The guy she likes

Her brother

You

Legend

COMPUTER VISION AND
Brazil

That's it, Messi, I'm not your supporter for nothing...

É isso ai Messi, sempre fui teu fã e não é a toa.

Manipulation to claim support

Original

A Federal University ‘class’
We don’t want to pay for it!

Image repurposing
India

AND THEN I SAID ...

GIRL MALNUTRITION IS DUE TO
FIGURE CONSCIOUSNESS

Ab Kya Karien?
Congress Ka Naam
BJP Rakhte Hai..

MODI FANS

My bosses at the RSS
want Indian women to
stay in the kitchen.

I'm way ahead.
I don't care if they exist.

ANSWER EVERY QUESTION WITH 'CLEAN CHIT' AND
GUJARAT DEVELOPMENT

COMPUTER VISION AND
Deep Fakes as a Political Tool

2018: BJP supporters target investigative journalist Rana Ayyub with a doctored pornographic video containing her face

https://www.huffingtonpost.co.uk/entry/deepfake-porn_uk_5bf2c126e4b0f32bd58ba316
Allah is great! Allah is great! Allah is great! at Bawaslu [election supervisory agency] tonight. Allah is with us.
Case Study: 2019 Indonesian Elections
Indonesian Presidential Election Overview

Direct general election
Fourth presidential election
Rematch of 2014 (five year terms)
Joko Widodo re-elected with over 55% of the vote
The Candidates

Joko “Jokowi” Widodo
○ Incumbent
○ PDI-P candidate (center-left party)
○ Appealed to younger voters
○ “Man of the people”

Prabowo Subianto
○ Challenger
○ Gerindra candidate (right populist party)
○ Strongly Islamic
○ Military background (Ties to Suharto dictatorship)
Indonesian Election Aftermath

Prabowo claims a count fraud
Protests/Riots in Jakarta May 21-22, 8 dead

Indonesia restricts access to social media

Survey of the Meme-space

Millions of images collected during the course of the election

**Twitter**
- ~170,000 images from 14 different sources

**Instagram**
- ~1.9 million images from 20 different sources
The Data

Image sources identified by our partner in Indonesia, CekFakta

Images span from May 31st, 2018 to May 31st, 2019

Collected from 2 users and 12 popular hashtags (Twitter)
JANGAN SAMPAI TERJADI !!
Buzzer Accounts

Social media propaganda

Firms use fake accounts to generate “buzz”

Buzzers use upwards of 250 fake accounts each

Finding Trending Memes: Motif Mining
Objectives for an early warning system

- Real-time processing pipeline with content flagging
- Assess meme landscape with respect to political actors
- Allow humans to gauge the threat of violence
- Scale to the order of billions of images per day
- Match composite images to other composite and donor images
Motif Mining Pipeline

i) BuzzFeed, imgur
facebook, Instagram, twitter

ii) Image Scrapers
Image Set
Feature Extractor
Image Index

iii) Query Selection

iv) Spectral Clustering
Sparse Affinity
Query Selection: Detecting Edited Objects

Use meme images to search for related meme content

- Option 1: Randomly sample images from known buzzers
- Option 2: Detect new memes via the identification of edits

http://reveal-mklab.iti.gr/reveal/
Detecting Image Edits

Detecting Image Edits

Detecting Deep Fakes

Iakovidou et al. Content-aware detection of JPEG grid inconsistencies for intuitive image forensics. JVCIR, 2018
Adapting a provenance pipeline to motif mining

Add the ability to localize and match small objects

Add a spectral clustering stage
Motif Mining (Twitter Dataset)

- 174,328 images scraped from Twitter
- Image Index was built in 6.2 hours on a single GPU machine with 12 CPU cores
- 10,000 Random queries were performed for matrix generation in 6.5 hours
- **No deep learning was used**
Motif Mining (Twitter Dataset)

- 174328 total images
- 197 computed clusters
- Mean cluster size: 895 images
- Median cluster size: 156
- Minimum cluster size: 26
- Maximum cluster size: 36304
Automatic Motif Detection
Automatic Motif Detection
Automatic Motif Detection
Some Questions...
Q1: What capabilities do we still need for meme analysis?
Q2: What aspects of semantic analysis can improve the process?
Q3: What is a good process for data collection?
Q4: Who outside of the forensics community should we engage with?