UNIVERSITY OF CALIFORNIA

Los Angeles

Computer Vision to Analyze Protests in Social Media

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Science in Statistics

by

Alexander Chan

2020

ABSTRACT OF THE THESIS

Computer Vision to Analyze Protests in Social Media

by

Alexander Chan

Master of Science in Statistics

University of California, Los Angeles, 2020

Professor Yingnian Wu, Chair

Images are central to understanding protests and mass activism today for its impact in shaping public opinion.

Previously, analyzing protest images required human annotation, which is laborious and expensive. In the modern era of social media, an automated and systematic method is required to analyze the vast amounts of social media images.

In this thesis, I introduce a deep-learning computational framework to analyze protest images. This system comprises of (1) processing and parsing social media images from Twitter, (2) a model to identify common protest image characteristics, such as violence, fire, and police, models to (3) detect and (4) classify faces of protesters to understand demographics, and (5) an deduplication algorithm to identify the most shared images.

The thesis of Alexander Chan is approved.

Frederic R Paik Schoenberg

Zachary Steinert-Threkeld

Jungseock Joo

Yingnian Wu, Committee Chair

University of California, Los Angeles

2020

Thank you to:

My family for their continued support in my education

My college roommates

and

The UCHA Co-op for providing me a place and community to call home.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Thank you to my research advisors Professor Joo and Steinert-Threkeld and the Statistics Department for funding my education.

VITA

2012	Lung Cancer Radiation Therapy High School Intern. Roswell Park Cancer Institute
2013	Methanogenesis Research High School Intern. California State University, Fullerton
2014	Pan-Cancer Research Intern. UCLA Biochemistry.
2014	Colon Cancer Research Intern. University of Nebraska Medical Center
2015	Supercapacitor Research Intern. UCLA Materials Science and Engineering
2015	ML REU Research Intern. University of Houston, Downtown.
2016	Management Consulting Intern. ZS Associates
2017	Statistics B.S. UCLA
2017	Data Engineering Intern. American Express
2018	Software Engineering Intern. Amazon
2019	Infrastructure Data Science Intern. Facebook
2017–2019	Teaching Assistant & Special Reader for Upper Division (Monte Carlo, Experimental Design, Computational Statistics) & Graduate Courses (Survey of Methods in Modern Statistics). UCLA Statistics
2018–present	Graduate Student Researcher. UCLA Department of Public Policy and Communications
2020	Statistics M.S. UCLA

PUBLICATIONS

Steinert-Threlkeld, Zachary, Jungseock Joo, and Alexander Chan. 2020. How Violence Affects Protests. In *APSA Preprints*. doi: 10.33774/apsa-2019-bv6zd-v2

CHAPTER 1

Introduction

Images play an increasingly greater role in the swaying public opinion on social issues. The advent of social media has given rise to the images and videos as the dominant medium for sharing and reporting on social movements. Images are impactful, they are able convey power, magnitude, and emotion in ways text simply cannot. No words can illustrate the gravitas and power imbalance of the Black Lives Matter movement as the iconic "Taking a Stand in Baton Rouge," where police brutality on Black communities is highlighted by 2 police in complete riot gear forcibly arresting a Black woman in a dress. Images are also succinct, which are important with ever decreasing attention-spans [12].

Figure 1.1: Taking a Stand in Baton Rouge by Jonathan Bachman

Social media images in particular have several distinct advantages when studying protests and mass activism. In general, social media are: closer to the action and subject to less censorship than traditional news-media. While traditional news media may have a large team of reporters across the world that may respond to events quickly, there is still a delay between when an event and media from the event being published. More often than not, social media posts are published instantaneously at the scene by participants and bystanders of the protest. Traditional news-media may miss quick skirmishes that, occur spuriously, such as the split-second when the first shot was fired in the Hong Kong anti-government protests in 2019. Social media is also more resistant to censorship than traditional media. Due to its decentralized nature, social media users in many countries are able to post without prior approval of an editor who may be under government supervision. While censorship of social media does exist in certain countries, it is often done on a lesser extent than traditional news media.

Figure 1.2: First shot fired of the HK Protests

Previous work on studying social movements have been dominated by surveys of protesters after a protest [15, 14], or formal and quantitative models [10, 11, 13]. Social scientists have previously studied images, however, much of these analyses were done on few images that were manually annotated. This is prohibitively expensive in terms of cost and time due to the scale of social media images. Images that were manually annotated were manually selected, which can lead to potential biases, such as cherrypicking data. In order to study images on a large scale, an automated and systematic way of analyzing images is needed.

There has also been several works in social multimedia and computer vision that use visual recognition to answer questions in political science and communication. Researchers have analyzed the perceived personalities from politician photographs on social media, where an automated system was able to infer perceived social traits such as intelligence, honesty and competence [5].

In this paper, I demonstrate a novel protest image pipeline to analyze images on social media. The protest image analysis pipeline consists of 5 main parts:

1. Twitter Image and Metadata Collection
2. Protest Image Classifier
3. Facial Detection
4. Facial Demographic Classifier
5. Image Deduplication Algorithm

CHAPTER 2

Twitter Data Collection

The data used in this project is from Twitter. Twitter is a social media platform where more than 500 million tweets per day are posted. Of those 500 million tweets, our server extracts the tweets using Twitter's streaming API and saves 5 million tweets per day. For a given protest event, we filter for tweets that match the country, time period of the protest, and those that contain an image. Twitter's tweets are arranged in a JSON format. Furthermore, we extract the metadata from tweets and make them into tabular CSV format. The metadata fields are listed below in Table 2.1.

Table 2.1: Metadata Table.

Metadata Field	**Description**
created_at	UTC time when this Tweet was created.
text	The actual UTF-8 text of the Tweet.
id	The integer representation of the unique identifier for this Tweet.
source	Utility used to post the Tweet. (Web, Mobile, etc.)
user.id	The integer representation of the unique identifier for this User.
user.location	The user-defined location for this account's profile. Not necessarily a location, nor machine-parseable.
user.followers_count	The number of followers this account currently has.
user.friends_count	The number of users this account is following.
	Continued on next page

Table 2.1 – continued from previous page

Metadata Field	Description
user.name	The name of the user, as they've defined it.
user.screen_name	The screen name, handle, or alias that this user identifies themselves with. screen_names are unique but subject to change.
user.statuses_count	The number of Tweets (including retweets) issued by the user.
user.created_at	The UTC datetime that the user account was created on Twitter.
user.utc_offset	User's timezone
user.verified	Indicates that the user has a verified account.
user.lang	User's language
geo.type	For Tweets with exact location, the type of location (e.g. Point)
geo.coordinates	For Tweets with exact location, the coordinates of the point.
place.place_type	The type of location represented by this place. (e.g. City)
place.name	Short human-readable representation of the place's name. (e.g. Manhattan)
place.country_code	Shortened country code representing the country containing this place. (e.g. US)
place.bounding_box.coordinates	A bounding box of 4 coordinates which encloses this place.
lang	Indicates a BCP 47 language identifier corresponding to the machine-detected language of the Tweet text.

Continued on next page

Table 2.1 – continued from previous page

Metadata Field	Description
entities.media.media_url	If Tweet contains image, this field contains the URL

Finally, of tweets that contain images, I download the images using the link media_url field. These images are used in later in the pipeline.

CHAPTER 3

Protest Image Classification

3.1 Protest Dataset Construction

To create a protest image classifier, a protest image dataset was created to train the classifier. 40,764 images were collected from Twitter and annotated by users on Amazon Mechanical Turk. Of the 40,764 images, 11,659 images are protest images. Among these protest images, we needed ground truth data on visual attributes, such as signs, fire, law enforcement, children, groups, flags, etc.

To annotate the 11,659 protest images, 58,295 image pairs were randomly sampled and 10 workers compared each pair to indicate which image of the pair looks more violent. The Bradley-Terry model was used to estimate the score of perceived violence - differentiated by state violence and protestor violence. The list of annotated visual attributes is listed in Table 3.1.

More details can be found in Won, et. al's "Protest Activity Detection and Perceived Violence Estimation from Social Media Images"[16].

3.2 Protest Model

Convolutional Neural Networks (CNN) are used to train models to detect such visual attributes. The model was jointly trained for the 2 tasks - protest classification and visual attribute classification. Binary cross entropy loss was used to train the binary variables

Attribute	Description
Sign	A protestor holding a visual sign (on paper, panel, or wood)
Photo	A protestor is holding a sign containing a photograph of a person
Fire	There is fire or smoke in the scene
Law Enforcement	Police or military are present in the scene
Children	There are children in the scene
Group 20	There are roughly more than 20 people in the scene
Group 100	There are roughly more than 100 people in the scene
Flag	There are flags in the scene
Night	It is at night
Shout	One or more people shouting

Table 3.1: Protest Model: Visual Attributes and Descriptions

(protest and visual attributes).

$$L_{BCE} = -\frac{1}{N}\sum_{n=1}^{N}[y_n \log(p_n) + (1 - y_n)\log(1 - p_n)] \tag{3.1}$$

where p_n is the predicted label and y_n is the ground truth binary label for the n^{th} image respectively. MSE was used to train the continous violence variables.

$$L_{MSE} = -\frac{1}{N}\sum_{n=1}^{N}(y_n - p_n)^2 \tag{3.2}$$

where p is the predicted value and y is the ground truth value.

The model architecture is based on a ResNet 50 - which consists of 50 convolutional layers with batch normalization & ReLU layers. The architecture of the model is listed in Table 3.2.

<table>
<tr><th>Layer</th><th>Output Size</th><th colspan="3">Building Blocks</th></tr>
<tr><td>Conv 1</td><td>112 x 112</td><td colspan="3">7 x 7, 64, stride 2</td></tr>
<tr><td rowspan="2">Conv 2</td><td rowspan="2">56 x 56</td><td colspan="3">3 x 3 max pool, stride 2</td></tr>
<tr><td colspan="3">$\begin{bmatrix} 1 \times 1, & 64 \\ 3 \times 3 & 64 \\ 1 \times 1, & 256 \end{bmatrix} \times 3$</td></tr>
<tr><td>Conv 3</td><td>28 x 28</td><td colspan="3">$\begin{bmatrix} 1 \times 1, & 128 \\ 3 \times 3 & 128 \\ 1 \times 1, & 512 \end{bmatrix} \times 4$</td></tr>
<tr><td>Conv 4</td><td>14 x 14</td><td colspan="3">$\begin{bmatrix} 1 \times 1, & 256 \\ 3 \times 3 & 256 \\ 1 \times 1, & 1024 \end{bmatrix} \times 6$</td></tr>
<tr><td>Conv 5</td><td>7 x 7</td><td colspan="3">$\begin{bmatrix} 1 \times 1, & 512 \\ 3 \times 3 & 512 \\ 1 \times 1, & 2048 \end{bmatrix} \times 3$</td></tr>
<tr><td>Pooling</td><td>2048</td><td colspan="3">average pooling</td></tr>
<tr><td>Classification</td><td>14</td><td>1-d fc (protest)</td><td>3-d fc (violence)</td><td>10-d fc (visual attribute)</td></tr>
</table>

Table 3.2: Protest Model Architecture

CHAPTER 4

Facial Detection

One problem of particular interest of researchers is understanding the people behind the protests. One avenue of approaching this is finding faces in an image and analyzing them. In the first stage, we detect faces, and in the second stage, we classify faces.

To detect faces, we use a CNN detector in dlib. The CNN model consists of 8x downsampling blocks. Each 8x downsampling block consists of 5x5 convolution layers that does 2x downsampling, and 3x3 convolution layers that do not perform any downsampling, with relu and batch normalization. The MMOD loss function is used, also called the Max-Margin Object Detection[8].

Commonly when an object detector is trained, large amounts of positive and negative training windows are presented to a binary classifier. The binary classifier is trained on these windows (or regions) of images. Furthermore, the classifier needs to be trained on completely negative images, where there are no targets in the image at all. The MMOD optimizer runs through all windows and minimizes the number of missed detections and false alarms[9].

The superb optimization performed by the MMOD can be seen by comparing its performance with Faster RCNN's face detection performance[4]. The MMOD detector outperformed Faster RCNN's detector in terms of recall and number of false alarms, though MMOD detector was trained with only 4600 faces while Faster RCNN was trained with 159,424 faces[7].

CHAPTER 5

Facial Demographic Classifier

Once human faces from each protest image were detected, they were cropped and classification was performed on them. In this classification problem, we were particularly interested in understanding the age, race, and sex of protest constituents (protestors and law enforcement).

5.1 FairFace Dataset Construction

To train a model that included protestor race, we first needed a dataset with human faces with ground truth labels for their race. While there were some existing datasets on this, such as LFWA+ and UTK, they only contained few races and missing important ones (or grouped as other) for Latino, East Asian, South East Asian, South Asian, etc. Moreover, datasets were highly unbalanced, with White faces highly over-represented compared to the rest of the dataset. For example, LFWA+, CelebA, and Coco all consist more than 85% of White faces[6]. Using unbalanced datasets is problematic, as unbalanced datasets create models that will over-classify classes that are over-represented, and under-classify classes that are underrepresented. Unbalanced datasets lead to models that are biased towards overrepresented classes.

As a result, we constructed a balanced dataset called FairFace, images were extracted from Flickr's YFCC-100M dataset and annotated using Amazon Mechanical Turk. Each race consisted of no more than 20% of observations. In total, about 108K human faces were annotated for the race, age, and sex.

<table>
<tr><th>Layer</th><th>Output Size</th><th colspan="3">Building Blocks</th></tr>
<tr><td>Conv 1</td><td>112 x 112</td><td colspan="3">7 x 7, 64, stride 2</td></tr>
<tr><td rowspan="2">Conv 2</td><td rowspan="2">56 x 56</td><td colspan="3">3 x 3 max pool, stride 2</td></tr>
<tr><td colspan="3">$\begin{bmatrix} 3 \times 3 & 64 \\ 3 \times 3 & 64 \end{bmatrix} \times 3$</td></tr>
<tr><td>Conv 3</td><td>28 x 28</td><td colspan="3">$\begin{bmatrix} 3 \times 3 & 128 \\ 3 \times 3 & 128 \end{bmatrix} \times 4$</td></tr>
<tr><td>Conv 4</td><td>14 x 14</td><td colspan="3">$\begin{bmatrix} 3 \times 3 & 256 \\ 3 \times 3 & 256 \end{bmatrix} \times 6$</td></tr>
<tr><td>Conv 5</td><td>7 x 7</td><td colspan="3">$\begin{bmatrix} 3 \times 3 & 512 \\ 3 \times 3 & 512 \end{bmatrix} \times 3$</td></tr>
<tr><td>Pooling</td><td>512</td><td colspan="3">average pooling</td></tr>
<tr><td>Classification</td><td>13</td><td>7-d fc (Race)</td><td>9-d fc (Age)</td><td>2-d fc (Gender)</td></tr>
</table>

Table 5.1: Facial Demographic Model Architecture

5.2 FairFace Model

Preliminary models to classify human faces were also trained. The model trained was a variant of the ResNet-34 that was altered to provide racial, gender, and age classification. The model architecture can be seen below in Table 5.1. The races classified by the model are: White, Black, Latino, East Asian, South East Asian, Middle Eastern and South Asian. The age groups classified by the model are: 0-2, 3-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, and 70+. More details can be found in *FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age* [6].

Even with no hyperparameter tuning, the model trained on FairFace outperformed models trained on LFWA+ (Labeled Faces in the Wild) [3] and UTK Face [17].

CHAPTER 6

Image Deduplication Algorithm

Another research question that we wanted to answer from a public policy and communications research perspective was "What images are most commonly shared?" This is useful to identify images that are visually striking and persuasive in shaping public opinion. While counting the number of retweets were handled by Twitter's API, an image deduplication algorithm was needed to count duplicate images from posted from original (non-retweeted) Tweets.

To deduplicate images, we extracted 1,000 features from a pre-trained ResNet50 model [2]. Conventional image preprocessing methods for deep learning models were used. Each image was resized to 256 x 256 px. Then, a center-crop of 224 x 224 px was performed. Finally, the cropped images were normalized to the mean and standard deviation of the ImageNet dataset [1].

The 1,000 feature vector of each sample was normalized to unit norm. The L2 distance among the normalized data is computed, and images are considered duplicates if the distance is less than a threshold of 0.2. This threshold was chosen after observing the pairwise distances of numerous duplicate and non-duplicate images.

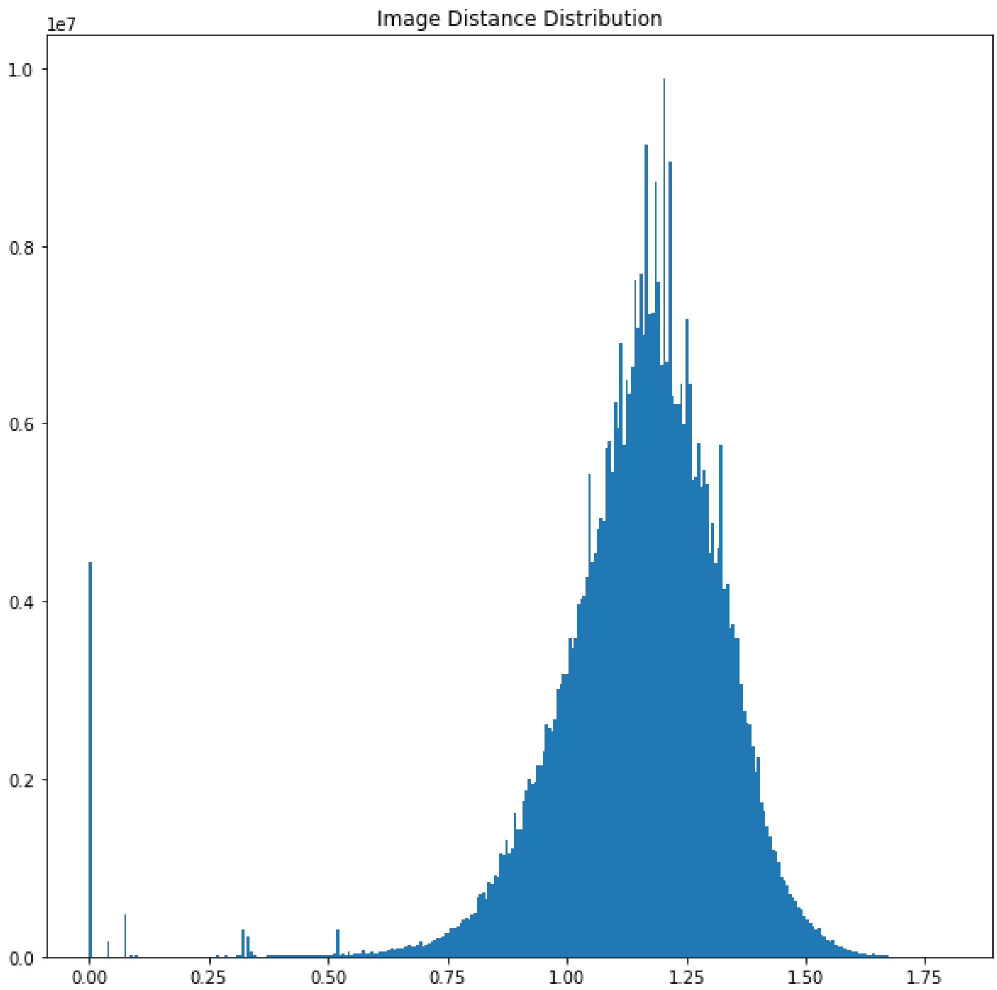


Figure 6.1: Pairwise distances histogram

CHAPTER 7

Case Study: 2019-2020 Hong Kong Protests

Starting in 2019 Summer, Hong Kong has been in a series of protests stemming from a controversial extradition bill. This bill would have allowed suspects to be extradited from Hong Kong, Special Administrative Region, to Mainland China, thus providing a legal conduit between Hong Kong's legal system with China's legal system. At the protest's peak, up to 2 million people, or about a quarter of Hong Kong's population, were protesting on Hong Kong's streets.

7.1 Protest Visual Attribute Classification

Although the protests are still ongoing, the protest periods of March 2019 to Dec 2019 were analyzed. The daily time series figures below show the proportion of protest images that a particular visual attribute is observed. A 10 day moving average was used as a smoother.

The Group 20 and Group 100 plots (Figure 7.1) show that a large proportion (40 to 80%) of the images show groups of at least 20, while a lesser, but non-trivial proportion (around 10%) of images show groups at least 100. This is because most images shot by protesters are up close, with a small field of view, to include groups of at least 20 protesters. While some images are far enough away to show at least 100 protesters.

We are also able to analyze the proportion of images with fire and flags (Figure 7.2). Flags were used throughout the HK protests. Some flags demonstrated a growing sentiment of anti-establishment, such as the Black Bauhinia flag, while some flags were an international plea for help. Some protesters also waved the UK Union Jack, reminiscent of Colonial Hong Kong under British rule. We observe that the proportion of flag images peaked in early

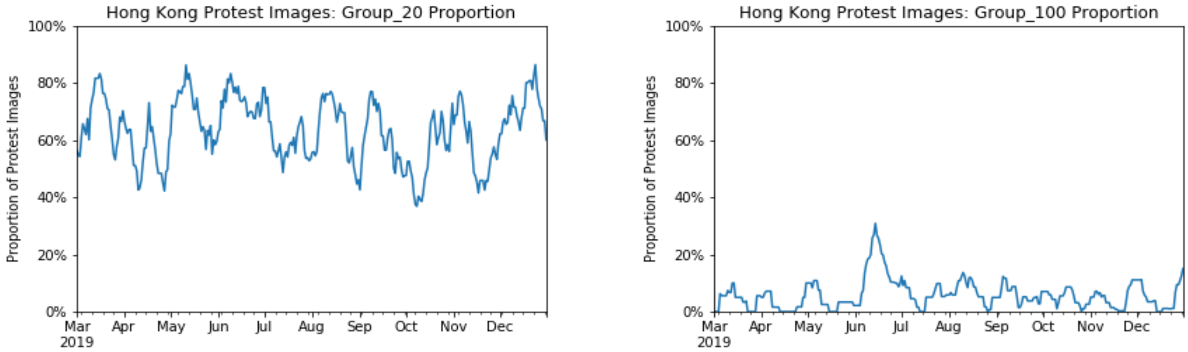


Figure 7.1: Time Series: Small and Large Groups

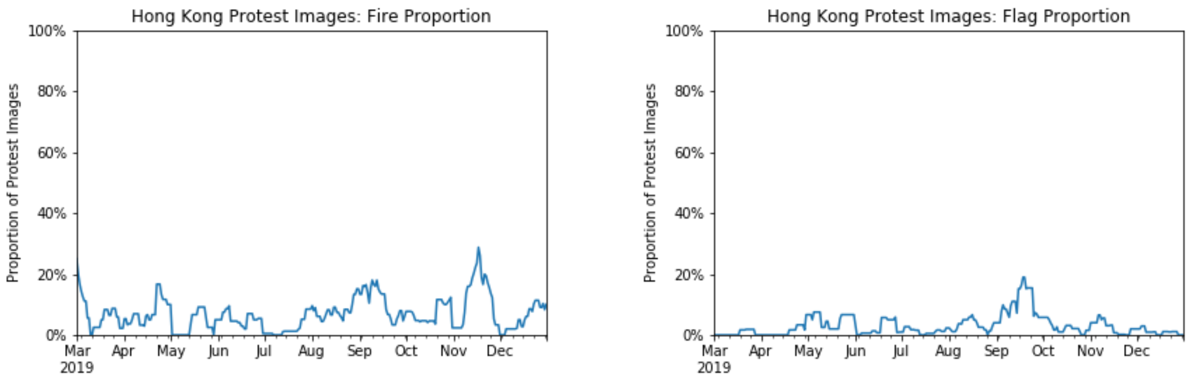


Figure 7.2: Time Series: Fire and Flags

September 2019. This coincides with the 8 September march, where thousands of protesters marched to the US Consulate to support for the US Congress's reintroduction of the Hong Kong Human Rights and Democracy Act. When the act was passed in US Congress in Nov 2019, 100,000 people rallied at Endinburgh Place where many protesters waived American flags. While the model was able to detect traditional flags well, it had trouble detecting non-traditional flags, such as the Black Bauhinia flag. Examples of these are shown in Table 7.1.

(a) Flag: 0.805285215

(b) Flag: 0.134031609

(c) Flag: 0.915045261

(d) Flag: 0.990652561

Table 7.1: Flags in 2019 HK Protests

The use of Fire was correlated with State and Protestor Violence. Fire was prevalent through the use of Molotov Cocktails and petrol bombs. In November 2019, the proportion of protest images that were observed with fire peaked around 28%. On 2nd November 2019, protesters at a rally in Central threw Molotov Cocktails on Lung Wo Road. During the

(a) Police: 0.822041035

(b) Police: 0.828614593

(c) Police: 0.911500454

(d) Police: 0.88702124

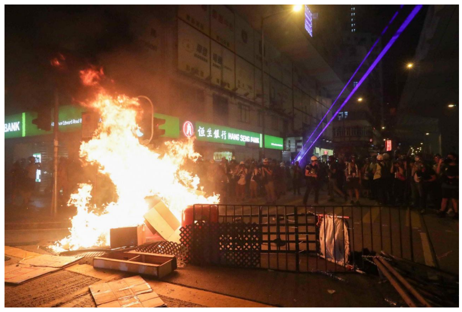
Table 7.2: Police in 2019 HK Protests

city-wide strike between November 11 and 15th, protestors threw molotov cocktails in a MTR train bound for Central. Tensions intensified, culminating in the siege of Hong Kong Polytechnic University from November 17th to 18th, where some protesters threw Molotov cocktails at the police as they attempted to raid the protesters at the university campus.

(a) Fire: 0.906967

(b) Fire: 0.842622

(c) Fire: 0.998881



(d) Fire: 0.936079

Table 7.3: Fire in 2019 HK Protests

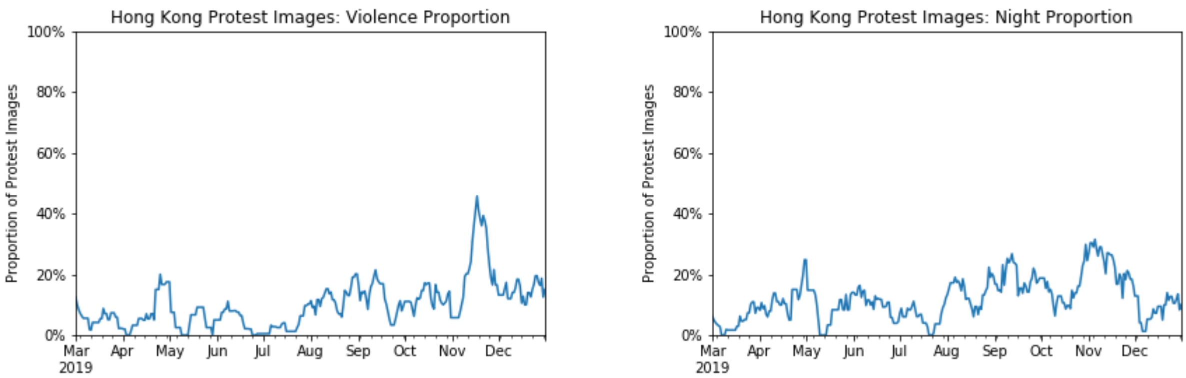


Figure 7.3: Time Series: Violence and Night

7.2 Facial Detection and Classification

We were also able to detect faces in protest images. Below is an example taken from a mass protest.

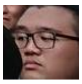
(a) SEAsian/M/20s

(b) E Asian/M/20s

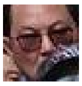
(c) NA

(d) NA

(e) NA



(f) E Asian/M/20s

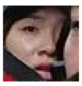


(g) NA

(h) E Asian/M/20s

(i) E Asian/M/20s

(j) E Asian/M/30s



(k) NA

(l) NA

(m) NA

(n) E Asian/F/20s

Table 7.4: Facial Detection and Classification

As we see above, the facial detection algorithm performs well for large faces, but does not perform well for small faces. Further, the facial classification algorithm requires faces to be large. The detected faces marked NA did not meet the facial classification size-threshold of 50 px $\times$ 50 px. Age ranges 20s means 20-29, and 30s means 30-39, etc.

7.3 Deduplication

Using the deduplication algorithm, we are able to see which photo has been Tweeted the most on Twitter.

The most Tweeted photo was Tweeted 46 times, show below in Figure 7.4. This is a scene during the Siege of Hong Kong Polytechnic University, where riot police raided a stronghold of protesters in the university.

Figure 7.4: Tweeted 46 Times

The 2nd most Tweeted photo was Tweeted 26 times, show below in Figure 7.5. This is another scene at the end of the Siege of Hong Kong Polytechnic University. Police arrested everyone who came out of the university, including medical volunteers and journalists as seen here.

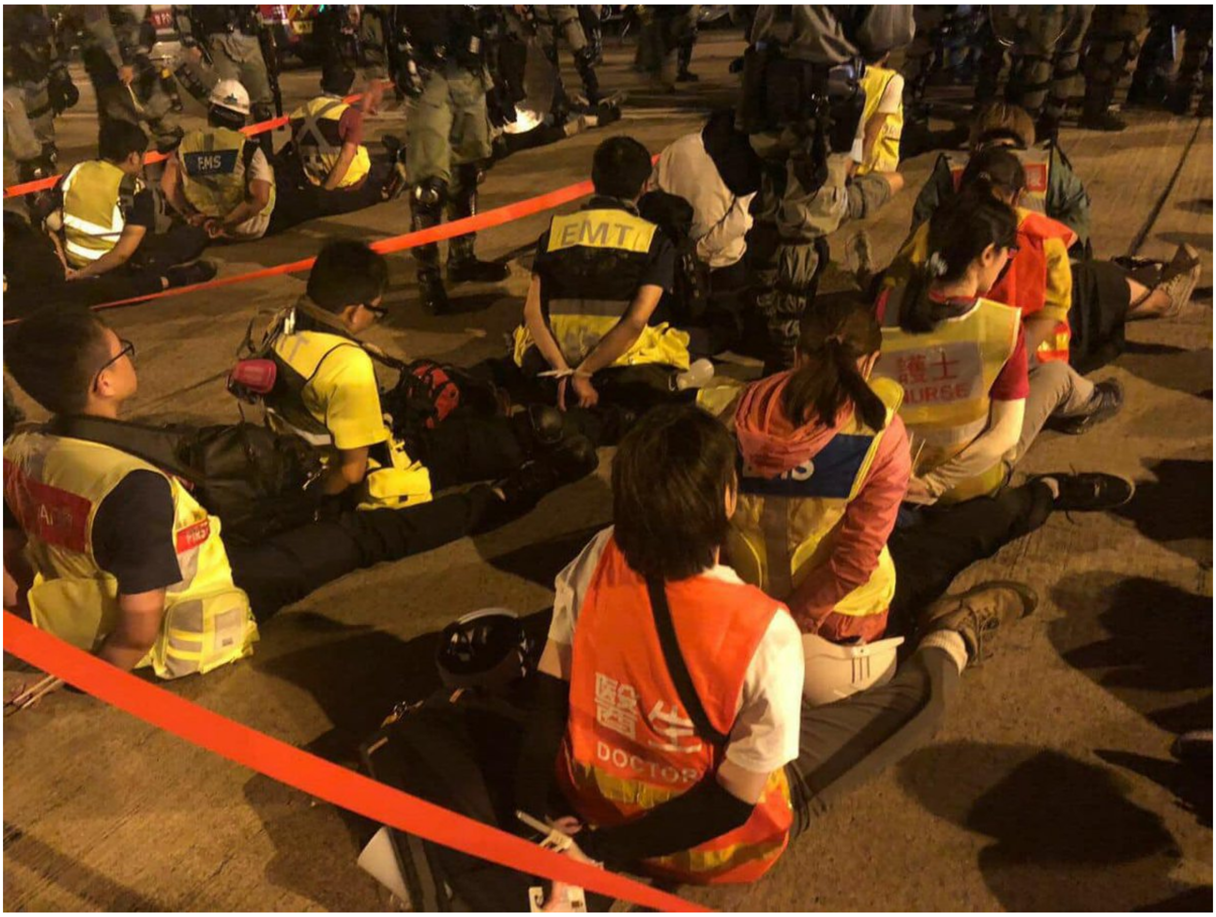


Figure 7.5: Tweeted 26 Times

The 3rd most Tweeted photo was Tweeted 14 times, show below in Figure 7.6. This photo caused an outcry where riot police pointed a shotgun directly towards a protester's head.

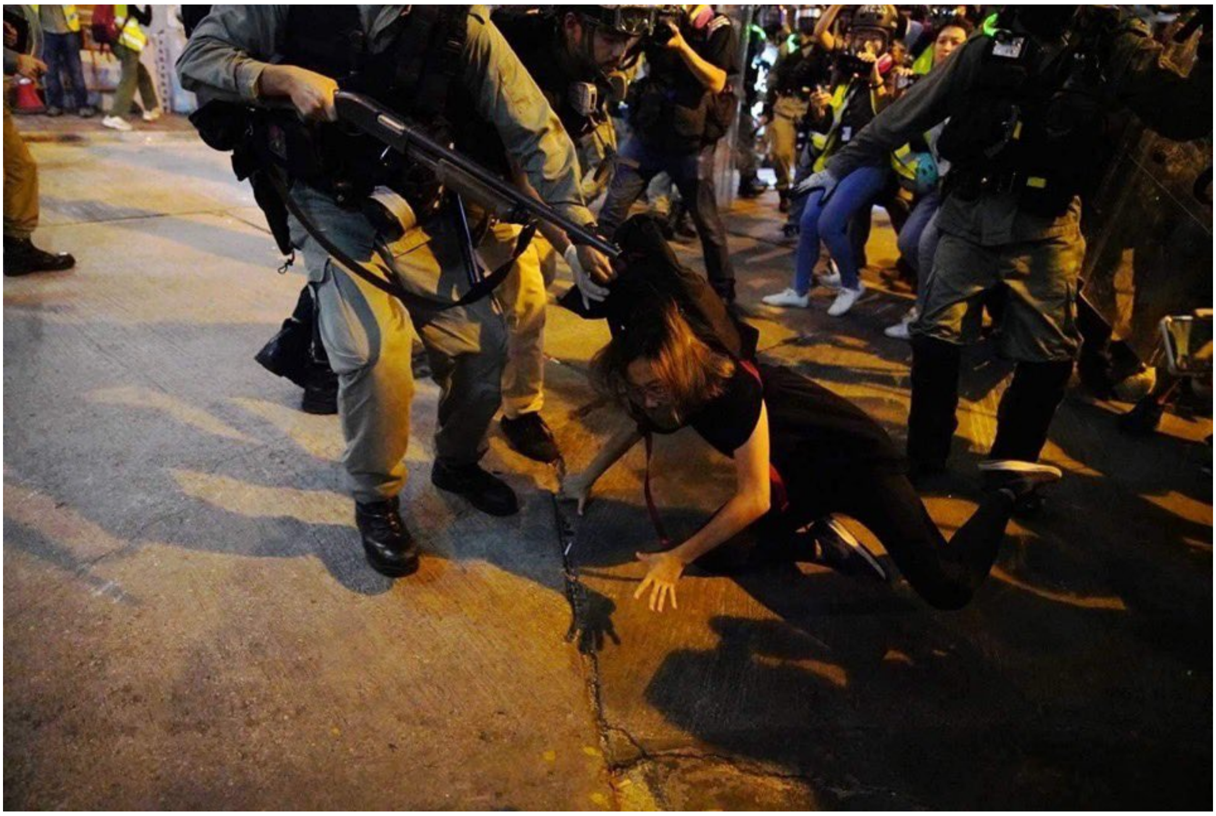


Figure 7.6: Tweeted 14 Times

The 4th most Tweeted photo was Tweeted 13 times, show below in Figure 7.7. This was another photo taken during the Siege of Polytechnic University. People were outraged after seeing a protester's head being stepped on by riot police.

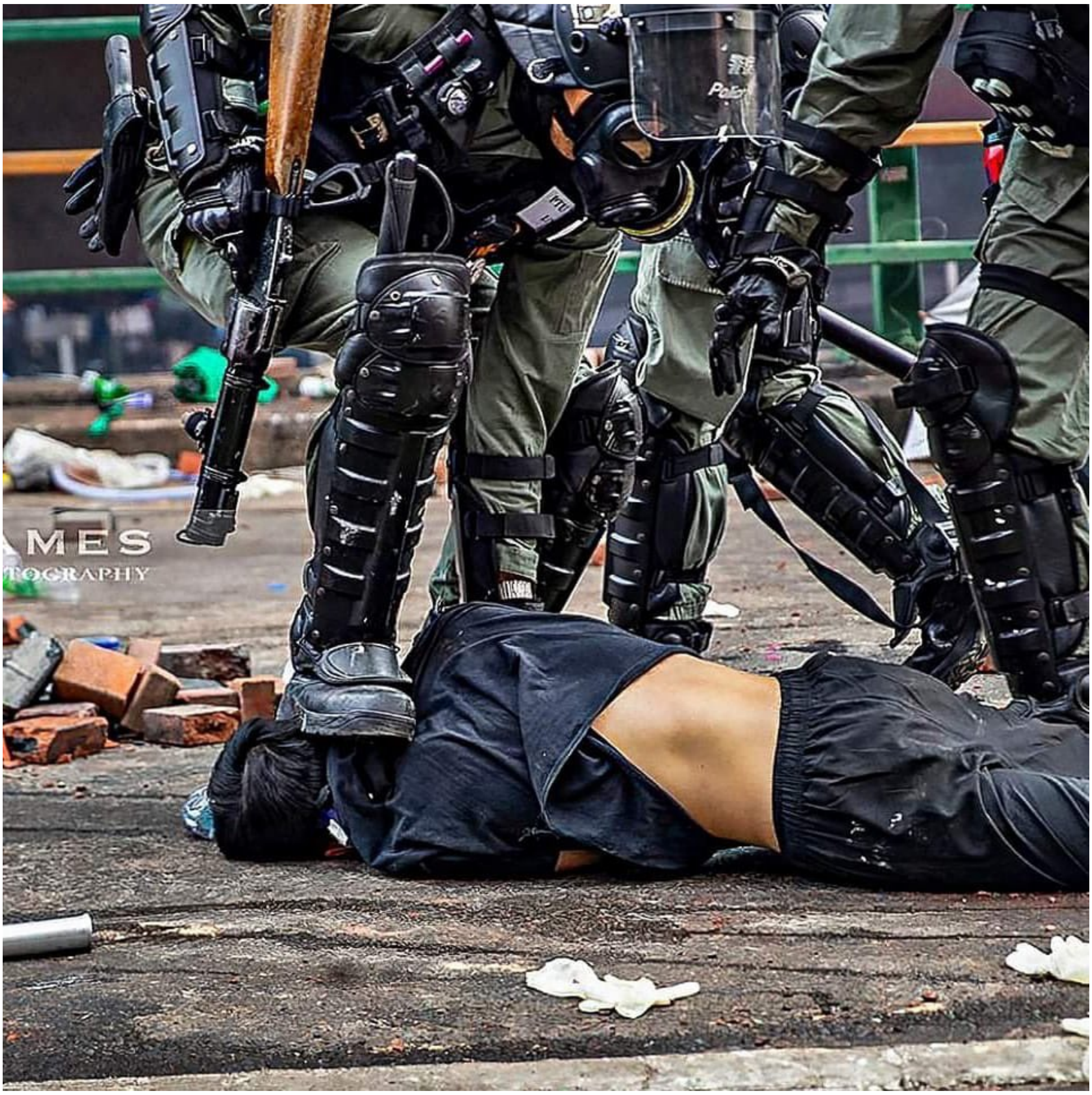


Figure 7.7: Tweeted 13 Times

The 5th most Tweeted photo was Tweeted 9 times, show below in Figure 7.8. Again, this photo was taken at towards the end of the Siege of Polytechnic University. In this photo, a riot policeman points his weapon at a protester as he tries to escape arrest.

Figure 7.8: Tweeted 9 Times

CHAPTER 8

Discussion

The model pipeline demonstrated so far can reasonably detect protest images and classify faces. However, there are some aspects where the computational framework could use some improvement.

The protest model sometimes has difficulty detecting non-traditional forms of protest (false negatives), and commonly identifies non-protests images as protests (false positives). The protest model also has difficulty identifying non-traditional flags, such as the Black Bauhinia flag. Since the model was trained on commonly colored flags, it missed blag flags. A more diverse training dataset and hyperparameter tuning can help to improve the precision and accuracy of the protest model.

The facial detection model performs well. However, it is limited by how it is trained. DLib trained the model with faces of size 50 px $\times$ 50 px or larger. As a result, the model struggles to identify smaller faces. Since the facial classification model is also trained on faces larger than 50 px $\times$ 50 px, it also struggles to classifies small faces. Thus, when trying to identify and classify faces from a large group of protesters, the model is accurate only with the participants close to the camera, which are larger. We particularly saw this in the facial detection example in the HK protests. A possible method to improve this is to train both models with smaller faces.

The deduplication algorithm is functioning, however it suffers from two main drawbacks: (1) the algorithm's time and memory complexity, and (2) misses from slight variations among visually identical images. The deduplication algorithm is reliant on a $O(N^2)$ distance calculation matrix. Likewise it the memory complexity is $O(N^2)$. While the distance calculation implementation has been modified to perform calculations in batches to avoid running out of memory, it is still not very feasible to run the deduplication algorithm on very large image-sets. As a result, when using this algorithm, we have only selected images that are protest images. However, many common images shared on social media related to a protest are not typical protesting images (e.g. large groups, police, weapons, fire etc.). These common images can be a banner or poster as a rallying cry. As a result, when using the deduplication algorithm in this manner, it will miss these images.

REFERENCES

[1] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[3] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. 2008.

[4] Huaizu Jiang and Erik Learned-Miller. Face detection with the faster r-cnn. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pages 650–657. IEEE, 2017.

[5] Jungseock Joo, Francis Steen, and Song-Chun Zhu. Automated facial trait judgment and election outcome prediction: Social dimensions of face. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 3712–3720, Dec 2015.

[6] Kimmo Kärkkäinen and Jungseock Joo. Fairface: Face attribute dataset for balanced race, gender, and age. *arXiv preprint arXiv:1908.04913*, 2019.

[7] Davis King. Easily create high quality object detectors with deep learning, 2016.

[8] Davis King. Dnn mmod source code, 2017.

[9] Davis E King. Max-margin object detection. *arXiv preprint arXiv:1502.00046*, 2015.

[10] Timur Kuran. Sparks and prairie fires: A theory of unanticipated political revolution. *Public choice*, 61(1):41–74, 1989.

[11] Susanne Lohmann. The dynamics of informational cascades: The monday demonstrations in leipzig, east germany, 1989–91. *World politics*, 47(1):42–101, 1994.

[12] Philipp Lorenz-Spreen, Bjarke Mørch Mønsted, Philipp Hövel, and Sune Lehmann. Accelerating dynamics of collective attention. *Nature communications*, 10(1):1–9, 2019.

[13] Doug McAdam. Recruitment to high-risk activism: The case of freedom summer. *American journal of sociology*, 92(1):64–90, 1986.

[14] Edward N Muller and Karl-Dieter Opp. Rational choice and rebellious collective action. *American Political Science Review*, 80(2):471–487, 1986.

[15] Zeynep Tufekci. Big questions for social media big data: Representativeness, validity and other methodological pitfalls. In *Eighth International AAAI Conference on Weblogs and Social Media*, 2014.

[16] Donghyeon Won, Zachary C Steinert-Threlkeld, and Jungseock Joo. Protest activity detection and perceived violence estimation from social media images. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 786–794, 2017.

[17] Song Yang Zhang, Zhifei and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017.