Protest Activity Detection and Perceived Violence Estimation from Social Media Images

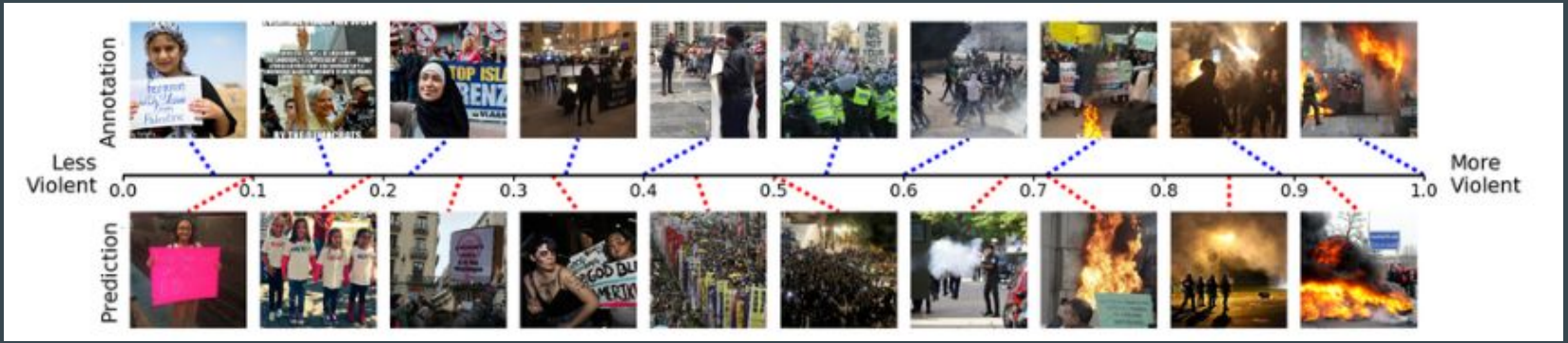
Donghyeon Won
UCLA
dh.won@ucla.edu

Zachary C. Steinert-Threlkeld
UCLA
zst@luskin.ucla.edu

Jungseock Joo
UCLA
jjoo@comm.ucla.edu

Introduction

- Great impact of social media on protests
- Limited by formal and quantitative models (protestor survey, etc.)
- More images are being shared but scholars have yet to analyze what they show
- Focus on violence in protests



Related work

- Little work in the fields of political science and media studies which attempts to automatically analyze visual or multimodal data due to the lack of proper methods and datasets
- Large-scale visual content analysis to tackle related research questions in political science, media studies and communication focus on facial attribute classification (age, gender, race)
- Public opinion about politicians has been also studied in relation to visual portrayals and persuasions in mass media

Dataset

- 40764 images
- Geotagged tweets
- 11,659 images are protest images identified by annotators and the rest are hard-negative images (e.g., crowd in stadium).
- This paper analyzes and compares five protest events including Black Lives Matter and Women's March.

Dataset

- The model aims to distinguish between a protest crowd and other large gathering such as concerts or sporting events. It should also distinguish between non-violent and violent protests.
- Trained rough Convolutional neural network based on positives examples and negative examples (concerts, stadium, flash mobs)
- Used Amazon Mechanical Turk (AMT) to obtain necessary annotations for each image in the dataset

Dataset


Annotation
Prediction
fire (1.00)
group>20 (0.64)
law enf. (0.50)
sign (0.99)
I'M 64!
I can't believe
I still have
to fight this
fight! Stop the
sign (0.95)
group>20 (0.86)
flag (0.64)
group>100 (0.53)
shouting (0.12)
fire (1.00)
group>20 (0.61)
night (0.97)
law enf. (0.59)
Violent
Angry
Fear
Sad
Happy

Models

- CNN taking a full image as input
- Series of prediction scores as output. Including the binary image category(i.e., protest or non-protest) (1), visual attributes (10), and perceived violence and image sentiment (1+4)
- Jointly trained the model such that all parameters for 3 different tasks – protest classification, violence and sentiment estimation, and visual attribute classification – are updated jointly

Models

1. In addition, another CNN captures various facial attributes from images.
2. OpenFace for face models (developed for face recognition)
3. Celeb A facial attribute dataset to train the attribute model.
4. For each image, dlib's face detection and alignment crops the internal facial region to feed into the facial CNN model.

Results

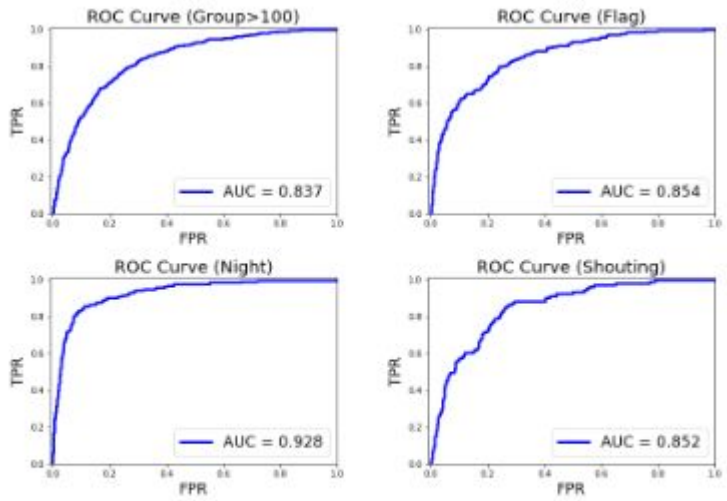


Figure 5: ROC curves for protest image and visual attribute classifications.

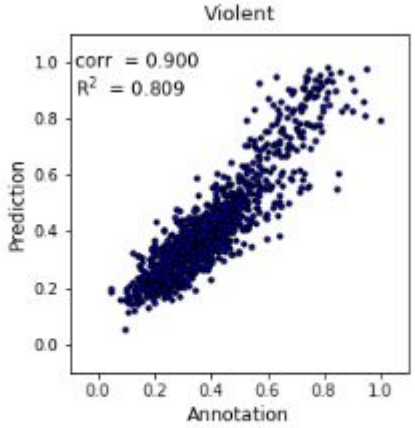


Figure 6: A scatter plot of perceived violence in annotations and predictions.

Results

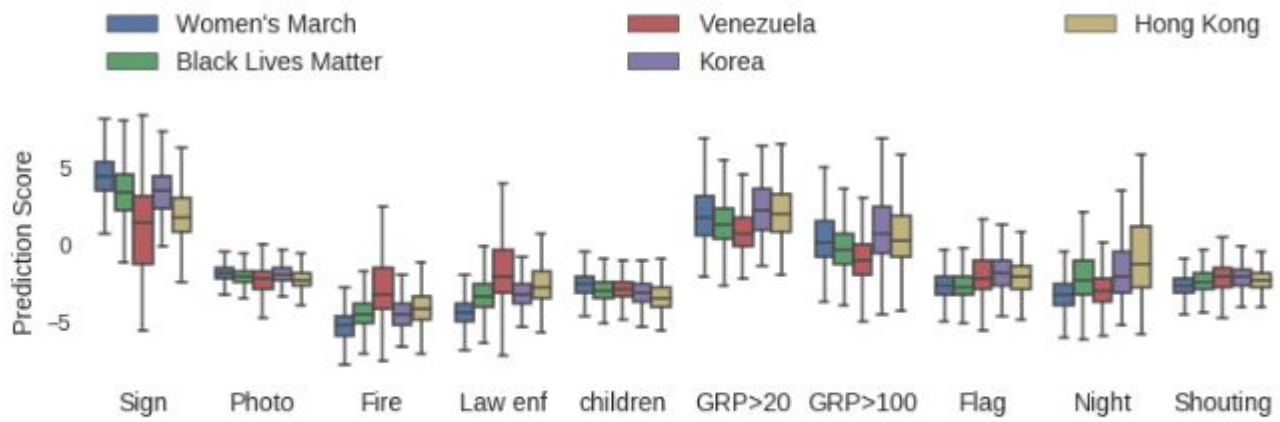


Figure 10: Predicted visual attributes in tweet images in different protest events. The box represents the range between the First and Third quartiles.

Conclusion

- New approaches to estimate violence and protest dynamics from social media images
- Primarily visual method of analysis
- Large training data set
- Collaborative area of research between multimedia and political science