

Data Appendix

Introduction

This project aims to utilize the methodology outlined in "Protest Activity Detection and Perceived Violence Estimation from Social Media Images" by Wondonghyeon et al. to analyze photos from the Hong Kong protests [1]. Through this analysis, we seek to gain insights into protest dynamics and the perceived levels of violence during these events.

Original Datasets Overview

- Entries: The dataset consists of 60 photographs during the Hong Kong protests.
- Variables:
 - The dataset contains the ImageName, protest, violence, sign, photo, fire, police, children, group 20, group 100, flag, night, and shouting features.
- Time Frame: The dataset was collected during the Hong Kong protests in 2019.
- Dataset CSV: The dataset was taken from the internet archive and the images were downloaded as a jpeg file [2].

DF Variable Descriptions

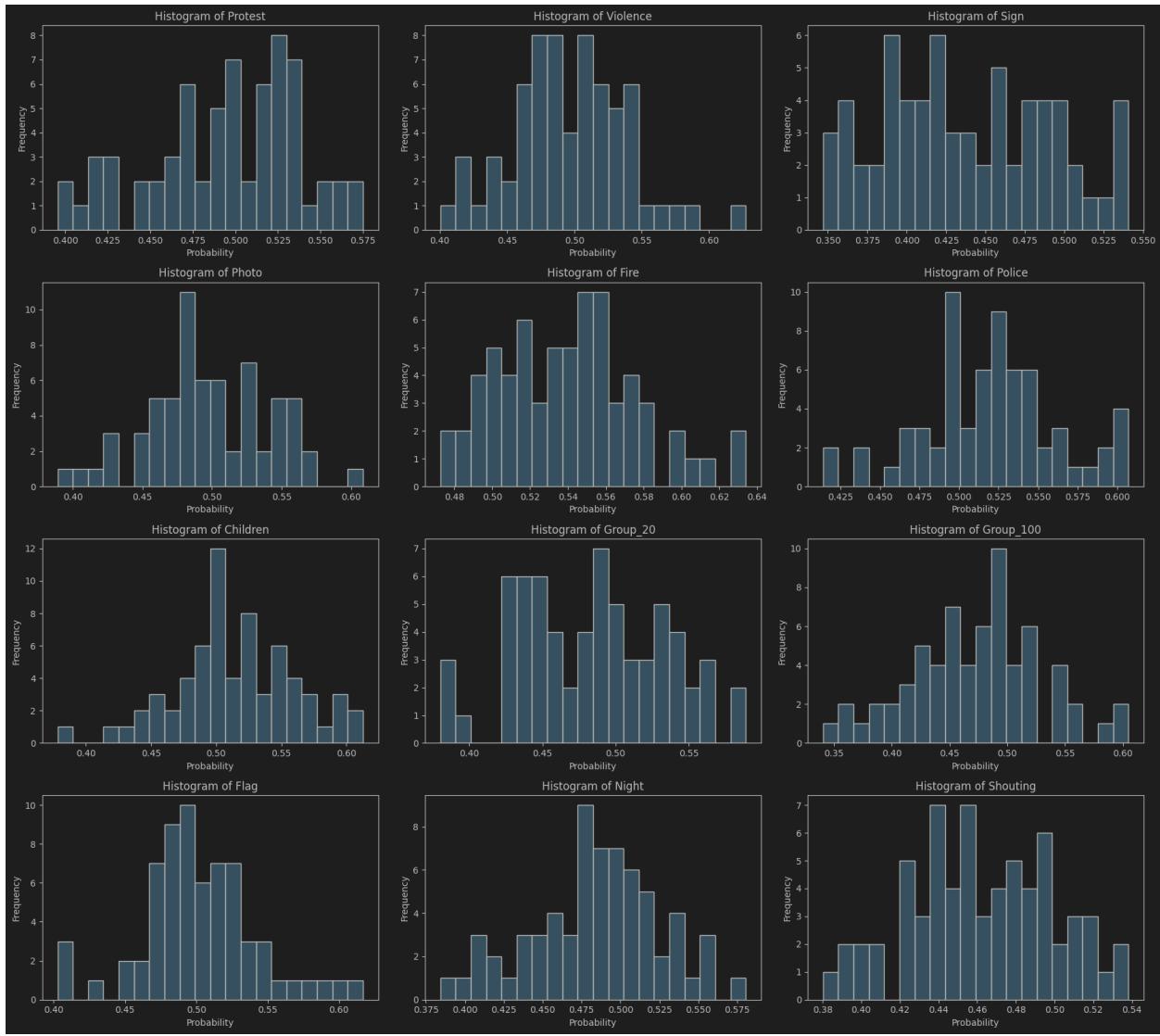
Data Dictionary:

Note: These are features so there was no variable type attributed to the data. The variable types below are in reference to the probability results.

Variable Name	Variable Type	Description
ImageName	pandas Series / String	The name of the image
Protest	pandas Series / Float64	The probability of a protest in the image
Violence	pandas Series / Float64	The probability of violence in the image
Sign	pandas Series / Float64	The probability of signs in the image
Photo	pandas Series / Float64	The probability of photos in the image
Fire	pandas Series / Float64	The probability of fire in the image
Police	pandas Series / Float64	The probability of police in the image

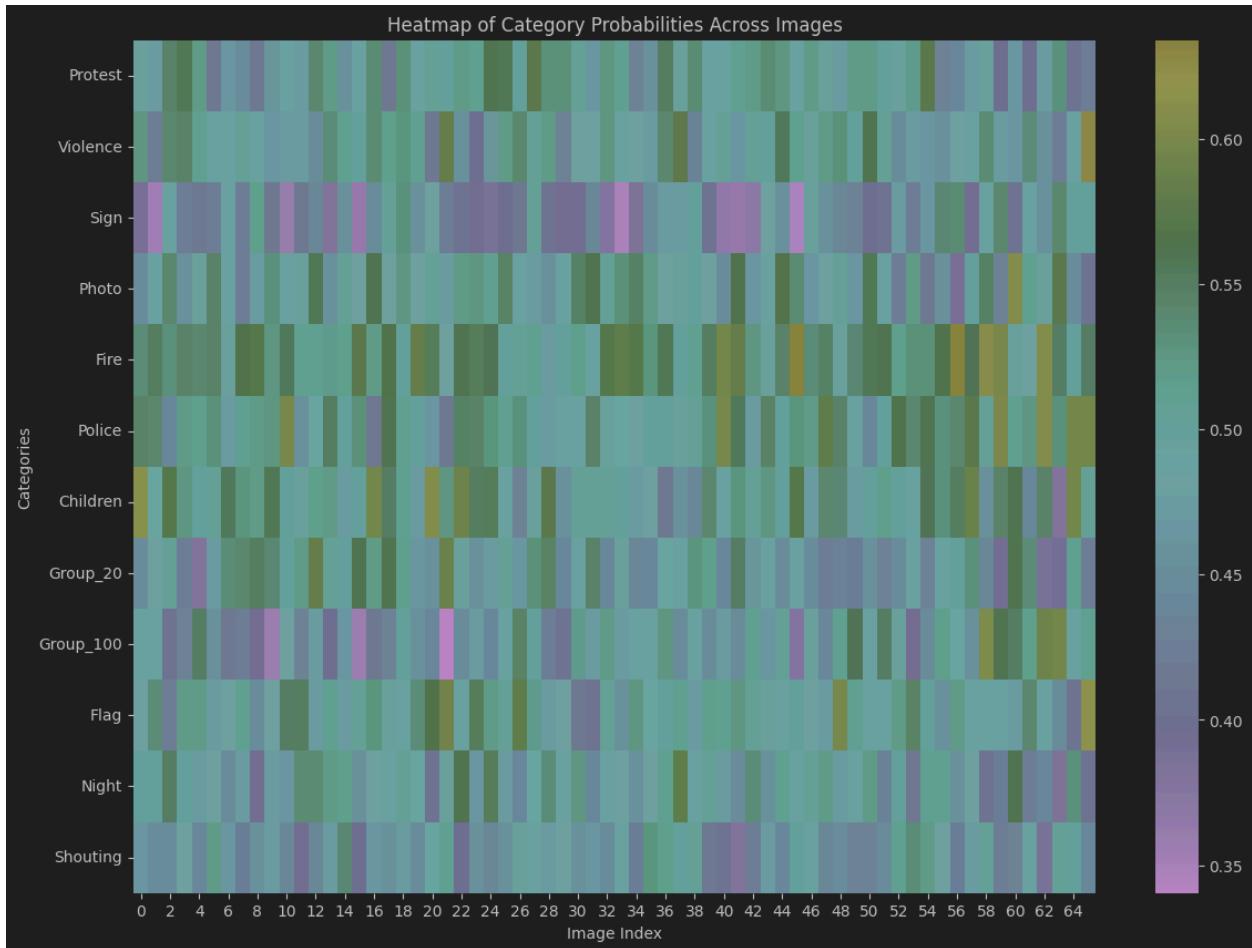
Children	pandas Series / Float64	The probability of children in the image
Group_20	pandas Series / Float64	The probability of there being more than 20 people in the image
Group_100	pandas Series / Float64	The probability of there being more than 100 people in the image
Flag	pandas Series / Float64	The probability of flags in the image
Night	pandas Series / Float64	The probability of it being night in the image
Shouting	pandas Series / Float64	The probability of shouting in the image

Detailed Statistical Analysis

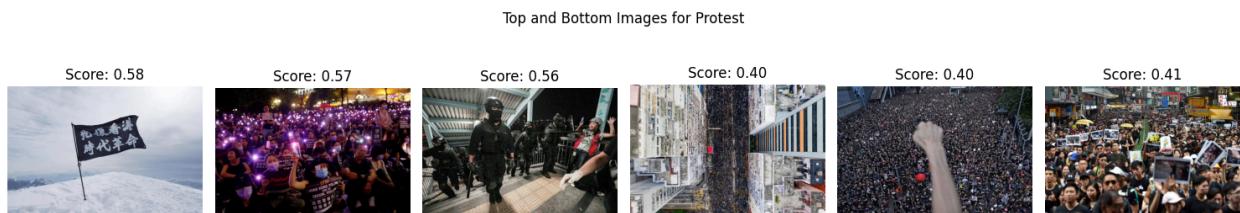


The provided graph exhibits a series of histograms representing the distribution of predicted probabilities for a range of factors associated with Hong Kong protest images, as determined by an AI model. Each histogram corresponds to a specific attribute such as Protest, Violence, Sign, and so on, showcasing the frequency at which certain probability scores occur within the dataset. Peaks within a histogram indicate commonality in the model's certainty regarding the presence of a particular attribute in numerous images. For instance, a peak in the 'Violence' histogram suggests a significant number of images were confidently identified by the model as depicting violence. Conversely, the broader spread in the 'Sign' histogram suggests a varied interpretation of signs presence in the images, indicating either a range of different signs captured or varying levels of model certainty. This distribution pattern allows for an understanding of which attributes are most predominantly and clearly represented in the dataset, according to the model's

assessment, providing insights into the nature of the protests and the model's interpretation capabilities.



The heatmap in the provided code visualizes the probabilities assigned to different categories across multiple images from a dataset. Each column in the heatmap corresponds to an individual image, while each row represents a different category. The cells of the heatmap are colored based on the probability values, with a color gradient (likely from green to purple, given the 'viridis' colormap) indicating the strength of the probability. Cells with darker colors represent higher probabilities, while lighter colors indicate lower probabilities.



Top and Bottom Images for Violence



Top and Bottom Images for Sign



Top and Bottom Images for Photo



Top and Bottom Images for Fire



Top and Bottom Images for Police



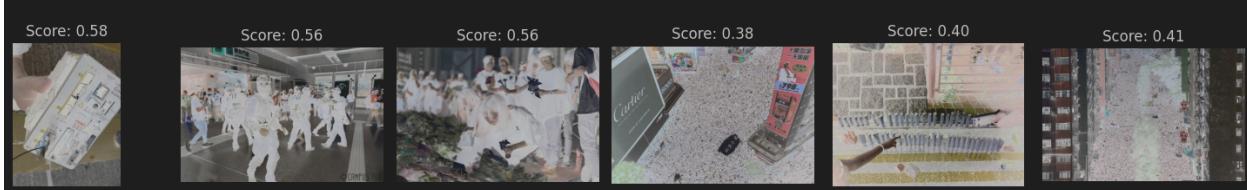
Top and Bottom Images for Group_20



Top and Bottom Images for Flag



Top and Bottom Images for Night



Top and Bottom Images for Shouting



Each of these graphs illustrates the three photos that scored the highest and lowest within each category. These visuals provide a comprehensive overview of the most and least favored images based on the respective criteria. By examining the top and bottom selections, viewers gain insight into the preferences and trends shaping the evaluation process. This comparative analysis aids in understanding the factors influencing audience perceptions and preferences, contributing to a deeper understanding of the data presented.

References

- [1] D. Won, Z. C. Steinert-Threlkeld, and J. Joo, "Protest Activity Detection and Perceived Violence Estimation from Social Media Images," in Proceedings of the 25th ACM International Conference on Multimedia, 2017.
- [2] "HongKongProtests2019_gallery_001," retrieved from Archive.org, [Online]. Available: https://archive.org/details/HongKongProtests2019_gallery_001/EC0SY18UEAA4HnU.jpg.