Machine Learning Object Detection for Open-Source Intelligence: Using Zero-Shot Object Detection & Natural Language Processing Text Classifiers to Identify Military Equipment

Author: CDT Leonardo Sabetta

Advisors: LTC Bruce Chojnacki & LTC Jessica Dawson

Introduction

The amount of information and multimedia contained on the Internet has grown exponentially in the advent of social media. The information surplus that has been created from this medium contains both official and unofficial forms of communication from celebrities, businesses, and governments at every level, as well as the United States military. Simultaneously, the rise of Machine Learning (ML) technologies and advancements in hardware and computing power have allowed this surplus of information to be filtered and analyzed on a large scale. Previously, this task would have been thought of as impossible without the most advanced systems available.

From time to time, the unauthorized sharing of information occurs, and while not classified, the aggregation of such data can be used to determine sensitive information. Indeed, the Defense Counterintelligence and Security Agency's term for this is "Classification by Compilation" [10]. Service Members (SM) are at particular risk of posting unofficial and unsanctioned images and video due to the widespread use of social media within the ranks. Posting military equipment in an unauthorized manner could provide context and insight to adversaries on the status and location of certain equipment. The aggregation of a given number of these posts, when tied to another factor such as time or location, could serve as confirmation as to what equipment the U.S. military has in a given region. The aggregation of this data and metadata may therefore reveal sensitive operational details such as troop movements or other potentially classified information. This presents an Open-Source Intelligence (OSINT) vulnerability as ML technology enables adversaries to filter through the information surplus and exploit the data for political or military gain. The current publicly available attempts to filter social media information have faced significant challenges. Accurate classifying of data is difficult and varies wildly due to factors such as camouflage, decoys or toy items, and varying environmental and lighting conditions. This study seeks to identify a method to reliably classify images of military equipment from social media networks, and in doing so, identify the proportion of virality for types of equipment as a first step towards constructing a tool that aggregates this data.

Literature Review

This study seeks to bridge the gap between existing ML methodologies used in natural disaster image classification and the specific challenges presented by military equipment classification to identify critical information that is available on social media and quantify the risks associated with this data, ultimately developing a framework that enhances both operational effectiveness and security in the face of emerging threats from social media data. The rapid development and prevalence of social media in the modern world has enabled its usage by a wide range of populations. Not immune to the addictive nature of social media [7], SMs across all branches have utilized this technology in unofficial and unsanctioned capacities that have enabled a vast trove of publicly available information to be accessed [9]. Artificial Intelligence (AI) is one such method that can be used to identify critical information. The methodology for classification of images through AI on social media has been conducted using various models. One such technique was using Transfer Learning (TL) in conjunction with a Convolutional Neural Network (CNN) with a small training set [1]. In Johnson's paper regarding the utilization of ML in the classification of natural disaster related images, he created a proof of concept that demonstrated the ability for a CNN with a comparatively small, labeled training set of 1,000 images to accurately identify hurricane related imagery. Utilizing this model, they were ultimately able to achieve 82 percent accuracy for the model across all the categorical classifiers. Natural Language Processing (NLP) Imran [2] is the process which enables the processing of associated textual information given an image (such as captions on social media) and thereby adds the ability for TL to be expanded to both the textual and image domain alike. Additionally, zero-shot models have been shown to effectively minimize the required manpower for the labeling of data, which is traditionally a large reason researchers tend to avoid large training sets [1]. Using TL via NLP in zero-shot models enables the use of an incredibly small manually labeled dataset, as well as crossover via the TL that enables a potential zero-shot model to classify broader categories if it can classify specific categories. Much of the pre-existing models that have been used at scale deal with the problem of classification of relevancy of data, with the most well-known usage of NLP being the Artificial Intelligence for Disaster Response (AIDR) system to classify whether a "tweet" is related to a natural disaster or not [1], [2]. The classification of military objects has long been troublesome for ML models, as the categorical classification accuracy in those images due to camouflage, decoys, as well as varying environmental conditions is poor.

The maximization of the security of a model and minimization of leakage can be obtained by using a local network with a dedicated computing station as opposed to the various implementations of artificial intelligence that require external networks and equipment. The implementation of localized artificial intelligence has been highly sought after, especially regarding the pharmaceutical sector [3]. This security risk is even more heightened in this project considering the potential for sensitive data involving troop movements or equipment to arise because of the aggregation of many sources of information. The aggregation of this information enabled by social media creates an OSINT vulnerability that can be accessed by anyone [4]. The lack of regulatory protection for the SM as well as the absence of regulation for posts that are non-political [8] in nature represent an unmanaged risk. The US military has allowed adversaries to exploit this unmanaged risk by seemingly ignoring this deficit in regulation. However, the aggregation of this information from the usage of social media by an opposing force presents a valid use case and would not mark the first instance of the usage of ML in military operations (as ML has been used to identify sea mines in real time) [5]. Therefore, a secondary goal of this research is to determine the feasibility of enhancing operational effectiveness via social media data-scrapping.

Research Methodology

The dataset this project uses consists of approximately 181 GB worth of video files from TikTok, the social media network, provided by ViralMoment, a for-profit company that specializes in analytics related to social media. Given the video quality provided, the dataset was 306 hours in total. ViralMoment enabled data collection from TikTok without regional bias, which allows the results of this study to be generalized across borders.

After obtaining the data, OpenCV's python library was used to extract every 30th frame from every video provided. A frame step rate of 30 was selected to maximize the efficiency of processing, as this step size is roughly 1 second at the video quality of the dataset. After extraction, the total resultant frame count was 1,099,951. Each frame was then provided to Microsoft's Florence2-large model, pretrained with FLD-5B (a dataset that contains 5.4 billion annotations across 126 million images), with a prompt of object detection. This model was selected as it has high zero-shot performance relative to other models, meaning it can skip a custom labeled dataset for this application. The text output of the objects detected were then tagged and associated with a frame using the frame step number and video ID as a unique identifier. After this, Facebook's Bart-large-MNLI, a zero-shot NLP model trained on the MultiNLI dataset (433k sentence pairs annotated with textual entailment information), was used to classify the text label into 1 of 5 categories: Aircraft, Ground Vehicles, Maritime, Weapons, Nonmilitary. A confidence threshold of 80% was used for each frame, with anything less than the threshold being classified as nonmilitary. Additionally, if more than half of the frames for a particular video were classified as nonmilitary, the category of the video was automatically classified as nonmilitary.

Initial attempts to validate accuracy resulted in a glaring error with text classification by Bart into categories, with any label that contained the string 'trousers' being classed as a weapon. This resulted in a second pass of all the text labels with OpenAI's GPT4 model. The OpenAI model identified 56,858 contested frames where the classification differed from Bart (a total of 5.1681% of the total resultant frames). A strict string filter was developed to supplement the classification process. Reasonable terms were generated for each category, and labels were checked against these terms. The string check was able to resolve 43,418 contested frames. Of the remaining 13,440 frames, there were only 738 unique frames. These frames were manually classified to their respective categories. Validation was performed against the Florence2-generated text labels by using a Mersenne Twister pseudorandom number generator to select a representative random sample of 100 frames from 100 different videos in each category. The labeling error rate across all 500 frames was 4.2% (21 frames). As evident through the validation process, the dominant systematic error occurs in the classification of the text labels, as the generated labels themselves did not show any patterned flaws.

Analysis

The results of the study found that 79% of the frames in the dataset were not related to the military. The next largest proportion was aircraft, at 10.55% of the total frames in the dataset, followed by ground vehicles at 7.77%. Maritime and weapons categories only made up 2.5% of the dataset combined. Key findings include the flaws of NLP when working with shorter length texts, as well as the virality of certain types of military equipment. As previously stated, the biggest disagreement between Bart and GPT4 was between the ground vehicles category, where over 30,000 frames were contested. The confusion matrix can be seen in Table 1 below. There were no identified consistent patterns in the erroneous Florence2 model text generation, and most of the object detection errors were due to imperfect angles and figures in the forefront obscuring clear line of sight to a given object. This is a well-known problem in object detection ML models. Additional Tables are available in Appendix A showing other related summary statistics.

Table 1. Confusion Matrix of OpenAI's GPT 4 Classifier Vs. Bart-MNLI

Final Data	**OpenAI**					
Bart-MNLI	*aircraft*	*ground_vehicles*	*maritime*	*nonmilitary*	*weapons*	**Grand Total**
aircraft	113,569	451	34	8	13	114,075
ground_vehicles	0	52,060	0	0	0	52,060
maritime	12	18	5,029	1,797	1	6,857
nonmilitary	2,240	32,545	6,526	865,521	1,543	908,375
weapons	6	26	3	11,635	6,914	18,584
Grand Total	115,827	85,100	11,592	878,961	8,471	1,099,951

Discussion

The findings of this study demonstrate that object detection has improved significantly, with models that are pre-trained on generic datasets are able to correctly and consistently label military vehicles such as tanks where they were unable to do so previously. A model with additional training could perform better but would require higher computing cost for the end user. These models could both be run on relatively lightweight hardware, with all data being processed on a device with only 1 RTX A6000. This level of computing is certainly available to state-sponsored threats, but also to lower-level belligerent groups that are funded through other activities. The level of ease with which this can be conducted demonstrates that an OSINT system that is able to perform this type of data processing at a much larger scale and connect this processing with metadata is not out of reach for the U.S.'s adversaries and competitors. The feasibility of creating a system that presents these risks should make the U.S. military consider the information disadvantage this creates when SMs post military equipment on social media. The potential for stricter control of social media usage by SMs is a valid concern, but the effectiveness of such a policy is also questionable given that the usage of TikTok on government systems was banned by the Department of Defense – yet there is still a superfluous amount of data provided by SM accounts.

This study finds that almost 90% of all military content posted on TikTok is split between aircraft and ground vehicles. These pieces of equipment are the most visible to the public and the most romanticized in popular culture and media (see *Top Gun*). It is expected that this trend would likely be similar on other social media networks, but this remains unproven. Future works can expand on this study by incorporating other social media networks, training the object detection model to classify labels of specific types of military equipment rather than broad categories, as well as simply utilizing a larger dataset.

Appendix A

Table 2. Manually Rectified Contested Frames (Duplicates Included)

user_categories	Count of Category
aircraft	260
ground_vehicles	406
maritime	28
nonmilitary	4,462
weapons	8,284
Grand Total	13,440

Table 3. Final Dataset Classification

Categories	Count of Category	Proportion of Total
aircraft	116,087	10.5538%
ground_vehicles	85,506	7.7736%
maritime	11,620	1.0564%
nonmilitary	869,983	79.0929%
weapons	16,755	1.5232%
Grand Total	1,099,951	100.0000%

Table 4. Final Military Dataset Classification

Categories	Count of Category	Proportion of Total
aircraft	116,087	50.4796%
ground_vehicles	85,506	37.1817%
maritime	11,620	5.0529%
weapons	16,755	7.2858%
Grand Total	229,968	100.0000%

Works Cited

[1]. Johnson, Matthew, et al. “Moving Emergency Response Forward: Leveraging Machine-Learning Classification of Disaster-Related Images Posted on Social Media.” *Journal of Management Information Systems*, vol. 40, no. 1, Taylor & Francis, Jan. 2023, pp. 163–82, https://doi.org/10.1080/07421222.2023.2172778.

[2]. Imran, Muhammad, et al. “Using AI and Social Media Multimodal Content for Disaster Response and Management: Opportunities, Challenges, and Future Directions.” *Information Processing & Management*, vol. 57, no. 5, Sept. 2020, p. 102261, https://doi.org/10.1016/j.ipm.2020.102261.

[3]. Meshkin, Hamed, et al. “Harnessing Large Language Models’ Zero-Shot and Few-Shot Learning Capabilities for Regulatory Research.” *Briefings in Bioinformatics*, vol. 25, no. 5, Oxford University Press, July 2024, https://doi.org/10.1093/bib/bbae354.

[4]. Hwang, Yong-Woon, et al. "Current Status and Security Trend of OSINT." *Wireless Communications & Mobile Computing (Online)*, vol. 2022, 2022, ProQuest, https://doi.org/10.1155/2022/1290129.

[5]. Munteanu, Dan, et al. “Sea Mine Detection Framework Using YOLO, SSD and EfficientDet Deep Learning Models.” *Sensors*, vol. 22, no. 23, Dec. 2022, p. 9536, https://doi.org/10.3390/s22239536.

[6]. Popov, Mykhailo, et al. “A Method for Object Classification in Aerial/Satellite Images with Incorporating Geospatial Information.” *Advances in Military Technology*, vol. 16, no. 2, Feb. 2022, pp. 309–31, https://doi.org/10.3849/aimt.01484.

[7]. Bhargava, Vikram R., and Manuel Velasquez. "Ethics of the Attention Economy: The Problem of Social Media Addiction." *Business Ethics Quarterly*, vol. 31, no. 3, 2021, pp. 321-359, ProQuest, https://doi.org/10.1017/beq.2020.32.

[8]. Sherman, Justin, et al. "Data Brokers and the Sale of Data on U.S. Military Personnel: Risks to Privacy, Safety, and National Security." 2023, techpolicy.sanford.duke.edu/wp-content/uploads/sites/4/2023/11/Sherman-et-al-2023-Data-Brokers-and-the-Sale-of-Data-on-US-Military-Personnel.pdf.

[9]. “Death by a Thousand Cuts: Commercial Data Risks to the Army.” *The Army Cyber Institute*, United States Military Academy, West Point, NY, n.d.

[10]. Center for Development of Security Excellence. *Derivative Classification Student Guide*. Defense Counterintelligence and Security Agency, 2021, www.cdse.edu/Portals/124/Documents/student-guides/IF103-guide.pdf.