Journal of Human Security | 2024 | Volume 20 | Issue 2 | Pages 9-15
DOI: https://doi.org/10.12924/johs2024.20202
ISSN: 1835–3800

Journal of Human Security

Research Article

Detection of Cyber Extremism Objects Through Utilizing Different Versions of YOLOv8 Model to Determine their Performance

Ala Berznji[1,2,3], Shvan A. Mohamed Amin[4]

[1]Computer Department, College of Science, University of Sulaimani, Sulaymaniyah 46001, Iraq
[2]Stockholm University, Department of Computer and System Sciences
[3]Cyber Security Department, IQ Group Holding, Sulaymaniyah 46001, Iraq. Email: alabe@dsv.su.se
[4]Cyber Security Department, IQ Group Holding, Sulaymaniyah 46001, Iraq. Email: Sharo.karwan@iq.email

***Correspondence:** alabe@dsv.su.se

Submitted: 19 August 2024 | Revised: 17 November 2024 | Accepted: 15 December 2024 | Published: 30 December 2024

Abstract: In this paper, we compare various YOLOv8 model variants on a custom-made dataset containing cyber extremism images to analyze their performance in detecting cyber-extremism content images. The experiment discussed was performed on 3 specific models: YOLOv8s, YOLOv8m, and YOLOv8x; the models were trained and assessed in the same conditions for a fair comparison. The images used for testing included visual elements that pertain to extremism, such as weapons, explosions, military vehicles, symbols, and ruined infrastructures. Manually collected a dataset of 1955 images containing 1126 labeled objects in 12 classes from open-source platforms. For the annotation of the TEGEMO dataset, the CVAT tool was employed, and bounding boxes were utilized for object detection and labeling. The models were trained on Google Colab of an A100 GPU, and performance was measured using standard object detection metrics like P, R, and mAP (mean Average Precision). According to the results, all YOLOv8 models were accurate in the detection of cyber extremism-related objects. Overall, YOLOv8x obtained the best mAP@50 score with 0.987, but YOLOv8s is the fastest in term of inference time, making it the compose to real-time applications. YOLOv8m was selected as the best compromise between accuracy and speed. The results validate the potential of the YOLOv8 models to aid the development of automated detection systems for the identification of extremist visual media. The study also discusses the limitations, including the small size of the dataset which varies in terms of extremist content, the problems of ethical considerations of using AI for surveillance, etc. Further work will involve the enrichment of datasets, preparation against adversarial attacks, and the employment of NLP for the discovery of violent content in a"multi-modal" context. This research serves as groundwork for the development of AI driven systems for combating cyber extremism, promoting cyberspace safety via intelligent visual content analysis.

Keywords: Cyber Extremism, Object Detection, Models, YOLOv8

In the digital age cyber extremism has become an alarming phenomenon because of the use of online platforms for everything from propaganda to recruitment and incitement to violence. The explosive expansion of social media and multimedia sharing drives an ever-growing challenge for law enforcement and technology providers to identify and contain extremist content in real-time. Traditional detection methods therefore suffer from scalability constraints, poor response time and difficulty in recognizing visual content embedded in images or videos. Deep Learning has seen substantial progress recently, especially in the field of Object Detection, and can provide new opportunities to automatically identify harmful content. Due to the high speed and accuracy of YOLO (You Only Look Once) models, they have achieved great performance in real-time object detection tasks. The most recent version was released by Ultralytics in October 2023, with architectural improvements like anchor-free detection and better data augmentation, as well as multi-tasking support like segmentation and classification. In this study, YOLOv8 is applied for the first time in the domain of cyber extremism object detection with a custom dataset that represents visual indicators, including weapons, explosions, military vehicles, and symbols that are commonly found in the extremist content. The research is conducted in such a way that it trains and evaluates three separate versions of the model (YOLOv8s, YOLOv8m, YOLOv8x) to find which version has the best trade-off between accuracy and processing speed. Filling a Vital Gap: Importantly, the work addresses a critical gap in current research which is focused on text-based detection, and thus most useful for practical applications in content moderation, surveillance, and digital forensics. It also notes challenges, including the importance of extensive datasets, labeling difficulty, and the potential ethical implications of AI-assisted monitoring of content, which underscore the need for responsible and secure use of detection technology.

The spread of destructive and violent ideologies through online [1] leads to one of the biggest problems in recent years, which is the detection of cyber extremism content. Using those data collection tools and analytics, tracking one of the manifestations of cyber extremism is quite a challenging task. When internet content came along, it became too hard for law enforcement and tech companies to see and stop the very extremist content that was spewing forth from such a place." Finally, the emerging and dynamic ways of cyber extremism where new propaganda and recruitment tactics are used will necessitate the use of a range of innovative and adaptive modern detection techniques. Image processing – has been a solving approach to many challenges e.g. object recognition, facial detection, image classification. [2]. Real-time object detection model, YOLOv8 is an advanced perception model that has showed its ability to identify and locate objects in images and videos [3] With four-fifths of a second processing time, its excellent perception capabilities allow it to process large volumes of visual data; hence, it is an ideal tool that can be utilized to detect the content that propagates cyber extremism. The major research works are performed applying image processing and YOLOv8 for the detection of cyber extremism content. YOLOv8 is also used to build a tool recognizing extremist content in [4] which obtains 96. 6% accuracy. In the same direction, [5] worked on detecting terrorist sources of propaganda using image-processing techniques such as object detection. Research regarding the probability of image processing and YOLOv8 usability in the detection of cyber extremism material. However, many challenges and caveats often accompany good outcomes. On the other hand, Despite the promising results, some challenges and limitations exist. And [6] studied the issue of data availability and quality, highlighting the significance of using extensive data sets used for both training and testing. The lack of labeled data and the difficulty obtaining realistic and diverse datasets makes the development of accurate models. Moreover, the ethical challenges of using AI in surveillance and monitoring were elaborated by [7]. However, the risk of biased algorithms and invasion of privacy makes these ethical issues must be handled carefully. Results YOLOv8 and image processing evaluated 54 products during the experiment. 55 to be significant [8].

Deep neural networks (DNNs) are robust, but they are also vulnerable to adversarial machine learning attacks. This novel image-processing method discovers adversarial photos to restore their true labels, allowing DNN-based systems to operate correctly once more.

Tests performed on an adversarial machine learning database confirmed its effectiveness, demonstrating its ability to generate varying types of attacks. [9]. The methods and procedures of NLP techniques help in analyzing extremist speech and detecting them existing on social media platforms [10]. Described discrimination techniques of extremism group language are being used by researchers to identify extremist ideology and stop its propagation [11]. Steganalysis techniques [12] can be used for examining images that contain hidden objects for secure communication among terrorist networks. Through text-based analysis and network and visual frameworks, this paper studies the context mining field on Extremism, Radicalization and Hate speech (ERH) domain.

The document has some very helpful guidelines for the various sectors of researcher's industries and governmental entities with a view to making cyberspace safer. The analysis uses NLP toolkits such as NLTK and GATE for keyword extraction and named entity recognition capabilities to identify potent keywords and named entities. The function of social network analysis is to understand how users and their aliases connect with one another. The analysis focusses on generating profiles from the pattern of the posts timings where authors can be traced with the advances of writing style fingerprints [13] which then can be used to identify single authors. The research paper studied manipulated image detection but showed difficulty between detecting anomalies in high-quality altered images as well as the detection of non-manipulated images incorrectly through false positive results of the algorithm [14]. Machine learning and deep learning-based image processing have reinvigorated the process of recognizing people doing activities on social media. The technology is used by different applications for manipulation detection and steganography systems as well as racism identification and the fake news detection respectively. Deep learning researchers applied two models for manipulated image detection, such as convolutional neural networks (CNN) in combination with generative adversarial networks (GAN) [15] This deep learning framework has addressed some challenges posed by real and fake images and their effectiveness in resolving the increasingly problematic issue of fake images on social networks [16]. Deep learning-based methods have transformed the detection of social media activity. Some of these techniques have been employed in different detection approaches and applied to the analysis of manipulation patterns, steganographic activity, racist messages, false news spreading, etc.

AI-based perception systems, such as object detection in autonomous vehicles [17], demonstrate the critical role of computer vision in safety-critical scenarios. Similarly, detecting extremist visual content in cyberspace demands high-speed and accurate visual recognition

Deep learning models such as CNN and GAN were used by researchers to detect manipulated images [18], the respective studies for face mask detection are detailed in Table 1 [19].

The primary objective of this paper is to most importantly collect a large set of images associated with cyber extremism, then a category is made into varied classes related to cyber extremism and accurately identify the item from in the classes. To this end, we leveraged three YOLOv8 variants models and trained them in accordance with the collected dataset and evaluated and contrasted those said variant models to identify the best performing in terms of detecting objects related to cyber extremism.

This document is therefore structured as follows: Section 1 presents the work. Related literature is presented in Section 2. In section 3 the results of models have been shown. Finally, section 4 synthesizes the main findings and suggests avenues for future research.

Table 1: Summary of existing research on detection of cyber extremism objects through utilizing different Models.

RefNo.	Year	Model	Description	Accuracy	Dataset used
[20]	2020	MDM (Mask Detection Model)	The model MDM was based on YOLOv3	93.00%	975 Custom dataset of masked with 850 unmasked
[21]	2020	YOLOv5	Using the YOLOv5 model, two classes were detected, and its performance was compared with other SOTA algorithms.	92.00%	Dataset of 9000 images. Source was private.
[22]	2020	SRCNet (Super-Resolution and classifier Network)	The SRCNet is an advanced deep learning framework specifically designed to accurately identify and categorize various states of mask usage, leveraging cutting-edge neural network architectures to ensure high precision and reliability in real-world applications.	98.70%	Total of 5035 images. Medical images Dataset contain: 671 images without mask, 1344 images of incorrectly worn mask and 3030 images of correctly worn masks.
[23]	2020	Inceptionv3	Inceptionv3, a 22-layer deep neural network developed by Google, is specifically trained to identify individuals who are not wearing masks, leveraging its advanced architecture for accurate detection.	99.90%	The research employed Simulated Mask Face Dataset.
[24]	2021	YOLOv2 and ResNet-50 were integrated into a single, unified model, combining their strengths for enhanced performance.	The research utilised YOLOv2 and Residual Networks (ResNets) to develop a robust system for detecting face masks.	81%	The model was trained using a combined dataset consisting of the Medical Mask Dataset and the Face Mask Dataset, both sourced from Kaggle.
[25]	2021	A Support Vector Machine (SVM) classifier was combined with MobileNetV2 to detect and classify images, utilising the lightweight and efficient architecture of MobileNetV2 for improved performance.	The MobileNetV2 model is lightweight due to its less dense convolutional neural network architecture, making it highly efficient for resource-constrained applications.	99.98%	Kaggle with Medical Mask dataset combined to train the model
[26]	2021	The model employed MobileNetV2, a convolutional neural network (CNN)-based architecture, in conjunction with OpenCV for enhanced functionality.	MobileNetV2 leverages frameworks like TensorFlow and Keras, along with OpenCV libraries, for efficient implementation and execution.	99%	Medical Masks dataset, available on Kaggle, was used for training and evaluation purposes.
[27]	2021	Deep Residual Learning employed for image Recognition	Two-stage detector model using Alexnet	98.20%	A manually curated and customised face mask dataset sourced from Kaggle was utilized for the study.
[28]	2021	YOLO v4 & YOLO v3	A custom dataset was developed, and various YOLO versions, including YOLO v1, YOLOv2, YOLOv3, and YOLOv4, were trained using this dataset to evaluate their performance.	71.69%	52635 images of novel dataset were proposed.

Table 1 (continued): Summary of existing research on detection of cyber extremism objects through utilizing different Models.

RefNo.	Year	Model	Description	Accuracy	Dataset used

[29]	2021	YOLOv4	The condition of wearing face mask was Identified by YOLOv4 model	98.90%	To train the model a total of 9599 images were used which contain: 7,959 images were sourced from the WIDER-FACE and MAFA datasets. Of these, 6,120 images were utilised for training the model, while the remaining 1,839 images were reserved for testing purposes.
[30]	2021	YOLOv4	YOLOv4 employs CSPDarknet53 as its backbone network architecture and incorporates the PANet (Path Aggregation Network) configuration at the neck level for enhanced feature extraction and aggregation. The head contains Neural network layers arranged in three segments.	99.50%	A total of 959 images were utilized, with 732 images allocated for training and 219 images reserved for testing the model.
[31]	2023	YOLOv8	The DS-YOLOv8 model, improved from the YOLOv8 model, addresses object detection challenges in remote sensing image tasks. It introduces a Deformable Convolution C2f module, lightweight Self-Calibrating Shuffle Attention,	97.7%, 92.9%, 89.7%	For this dataset, 2,170 images were randomly selected for the training set
[32]	2024	YOLOv8	The study compares YOLOv8 and Mask R-CNN machine learning models for instance segmentation in agriculture.	90.2%	Two datasets of 1553 RGB images from orchard lighting conditions were prepared for deep learning analysis. Dataset 1 included 474 dormant season images annotated manually, while dataset 2 included 1079 green fruit images.
[33]	2024		The study introduces a new approach to improve the accuracy of the YOLOv8 model in object detection, focusing on small objects.	97%	(COCO) dataset
This research	2024	YOLOv8	The model has been trained for detect cyber extreme content	98.50%	1482 images has been used for training with 12 different class custom dataset

I describe the research methodology which determines how YOLOv8 performs object detection when applied to cyber extremism cases. The paper details YOLOv8's architectural structure by exploring its version upgrades while describing its head-level and backbone sections. The method describes the methods for gathering data as well as annotation and training procedures that result in the creation of an effective model for detecting objects in complex visual environments. Additional details about the methodology are provided for complete comprehension of the research procedure.

2.2. Models Architecture

The architecture of YOLOv8 is a complete redesign compared to previous versions like YOLOv5. It adopts a three-stage end-to-end design consisting of a backbone, neck, and head. To extract hierarchical features while avoiding excessive computation, we use CSPDarknet as the backbone with C2f modules. It allows for deep feature reuse and smoother gradient flow, enabling learning across a more significant number of layers. The neck has been implemented using PANet, which combines low- and high-level information in many paths and enables the model to generate detections with different scales. This is especially advantageous when determining small or far away cyber extremist objects, such as drones or weapon debris. The head takes the anchor-based mechanism that has been in existence for such a long time, instead, it does anchor-free object prediction. This simplifies configuration and increases model robustness. YOLOv8 employs a unified detection layer that predicts object coordinates and class scores in one go. It jointly supports segmentation and classification tasks and is therefore adaptable to intricate visual environments. Auto-learning bounding box assignments, enhanced label smoothing, and dynamic input resizing during training are among the additional features enabled by the architecture. It also has built-in Augmentation strategies like mixup and mosaic for improved generalization. These upgrades enhance recognition in low-light, cluttered, or partially occluded scenes, all common in cyber extremism imagery. YOLOv8s256, YOLOv8m, and YOLOv8x, have varying layer depth and width, allowing for flexibility based on computational power and accuracy needs. Of these two models, YOLOv8s has faster inference time and the least params, YOLOv8m is a middle-of-the-road option, and YOLOv8x has the best accuracy with higher latency and memory usage. ONNX, TensorRT or directly on edge devices — All models can be deployed. Its modular architecture supports quick adaptation for new datasets and tasks, making YOLOv8 applicable use cases ranging from real-time cyber threat monitoring to law enforcement investigations and automated moderation systems.

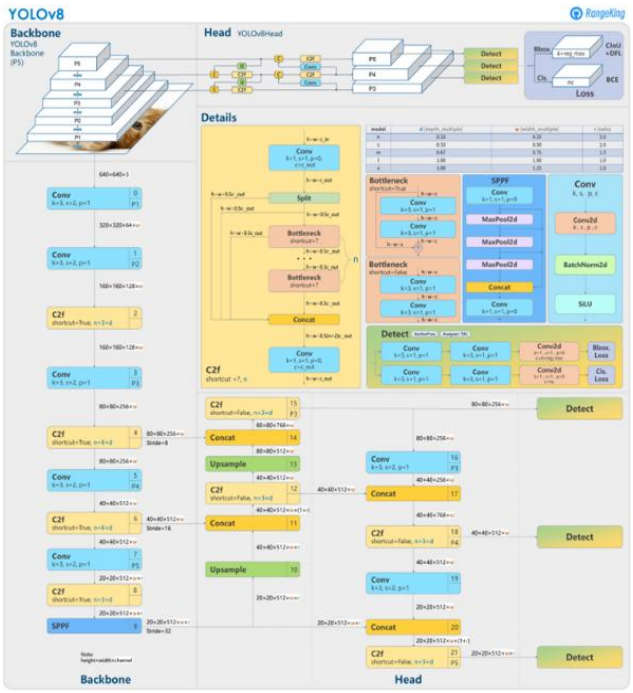

Figure 1: Architecture structure of YOLOv8.

2.3. Head

Re- https://towardsdatascience.com/3d-pose-and-morphing-in-3d-space-3d4e251019a To To[ˈrɛgrɛʃən] [eɪ/] put out or back to regular in contrast to conventional techniques, Regression-based approaches have become the main techniques to cope with the drawbacks of the standard approaches in 3D human pose estimation. This facilitates the supervised regression of these parameters as deep learning approaches establish direct mappings between the images and the associated human body pose and shapes. Large-scale training data and further improvement of neural networks have produced results in машинное поучение that are both tightly precise and dependable, according to scientists. The head section is the highest level of the network used by YLONv8 that operates in the extracted feature map space after the feature maps have been computed. The detection layers with up sampling layers and route layers are the vital components of the head segment. Predicted bounding boxes perform object detection tasks via detection layers feature maps.

Essentially, during its operation YOLOv8 detection layer encodes feature maps with a series of convolutions to produce bounding boxes with category prediction probabilities at multiple scales. By having an anchor box in each detection layer, different sizes of objects can be

detected. Up Sampling layers are used to remove layers and increase the resolution of feature maps. There are two functions performed by deconvolution operations in these layers: both recover input data and create high resolution output from low resolution input data. The up-sampling layers are the primary mechanisms that enhance the detection of small-sized objects. Network connections are formed through several route layers from the different levels of feature maps. This process links features maps of the previous layer to produce feature maps of multiple scaled information connected to feature maps of the earlier layer.

The model can detect objects of various sizes and specific qualities as feature pyramids are combined with various scales. In YOLOv8, box detection takes place in the "head" element through the detection layer as well as the up-sample layer and route layer, making the process very similar to the previous versions of YOLO. It also provides the coupling between multi-scale features to effectively detect targets of different scales and types.

2.4. YOLOv8

YOLOv8 introduces major improvements in both design and usability. It is not just an upgrade from YOLOv5, but a complete redesign focused on modularity, performance, and flexibility. One of the core architectural changes is the removal of anchor boxes. YOLOv8 adopts an anchor-free strategy where the model directly predicts object centers and bounding box coordinates without relying on preset sizes. This reduces the need for manual tuning and improves generalization, especially in datasets with diverse object scales like cyber extremism imagery.

YOLOv8 uses a modified backbone with C2f (Cross-Stage Partial with fusion) blocks, which enhance learning efficiency by reusing features and reducing redundancy. These blocks improve gradient propagation and stabilize training. Combined with depthwise separable convolutions, the model becomes more lightweight without losing accuracy. The neck section leverages PANet for effective multi-scale feature fusion, enabling detection of both large vehicles and small symbols in the same image.

On the training side, YOLOv8 includes built-in support for advanced augmentation like mosaic, mixup, and random perspective transformation. These techniques simulate real-world distortions, which help the model handle noisy or low-quality images. The training pipeline includes adaptive learning rate, cosine annealing, and warm restarts, which boost training efficiency and prevent overfitting.

Functionally, YOLOv8 supports multiple tasks:

- Object detection (bounding boxes + class prediction)
- Instance segmentation (pixel-level object outlines)
- Classification (image-level prediction)
- Pose estimation (human body keypoints detection)

Deployment is simplified. YOLOv8 supports export to TorchScript, ONNX, CoreML, and TensorRT for inference on various platforms including cloud servers, edge devices, and mobile apps. This flexibility makes it practical for real-time surveillance, content filtering, or law enforcement use.

YOLOv8 also includes performance tracking tools. It logs training metrics, validation accuracy, and loss curves automatically. It can resume training from checkpoints and supports mixed-precision training to reduce GPU memory use and increase speed.

The model sizes—YOLOv8s, YOLOv8m, YOLOv8l, YOLOv8x—scale the number of layers and parameters.

YOLOv8s is fast and ideal for edge deployment.

YOLOv8m balances accuracy and speed for general-purpose use.

YOLOv8x provides high precision for complex tasks but requires more memory.

For cyber extremism detection, these features are critical. The ability to detect small and obscure symbols, damaged vehicles, or weapons in complex environments requires robust spatial feature extraction and multi-scale prediction. YOLOv8 delivers this while maintaining inference times that support real-time analysis.

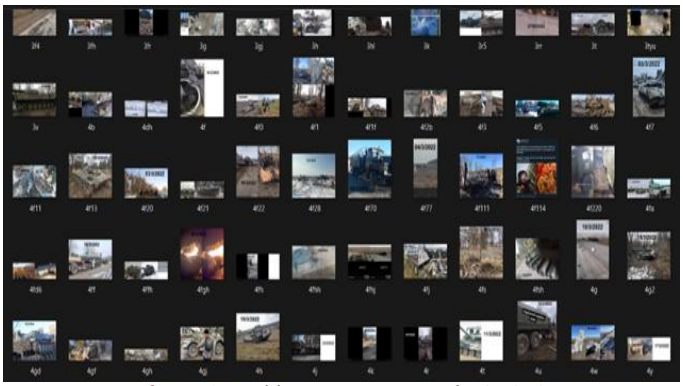

Figure 2: Architecture structure of YOLOv8

2.5. Data Annotations

The data annotation involved in this research was intensive and detailed because it had a direct influence on the ability of the model to learn useful patterns in terms of cyber extremism. The 1955 collected images were manually annotated using CVAT, and for each image annotators drew bounding boxes around any visible object that appeared to be relevant to extremist behaviors and environments. The visual analysis involved identifying things like military hardware, arms, wreckage, symbols used in propaganda, or other recognizable hallmarks of areas in conflict. The 12 distinct classes allowed the model to differentiate between functionally similar objects that might be visually similar. The annotation was in the YOLO format where each label consisted of the class index and normalized bounding box coordinates. To reduce human error and ensure improved consistency, the team embraced a methodical workflow that involved annotation, subsequent double-checking, and peer verification. Annotators relied on a pre-established guideline for the labeling in order to avoid subjective judgments and ensure consistency across the dataset. Dividing images into parts resulted in better clarity, particularly in scenes with overlapping or partially hidden objects. These variations (lighting conditions, object sizes, image qualities, background noise, etc.) were intentionally included to mimic the diversity of real-world extremist content on the Internet. As the dataset needed to be constructed from the ground up, the annotation phase took significant time but was critical to training a working object detection model. An exceptional attention to detail ensured that YOLOv8 would then train on accurate, consistent, relevant input, resulting in high precision and recall during testing. Ultimately The robust and flexible detection of complex visual signifiers of extremism was achieved has through the thorough annotation strategy.

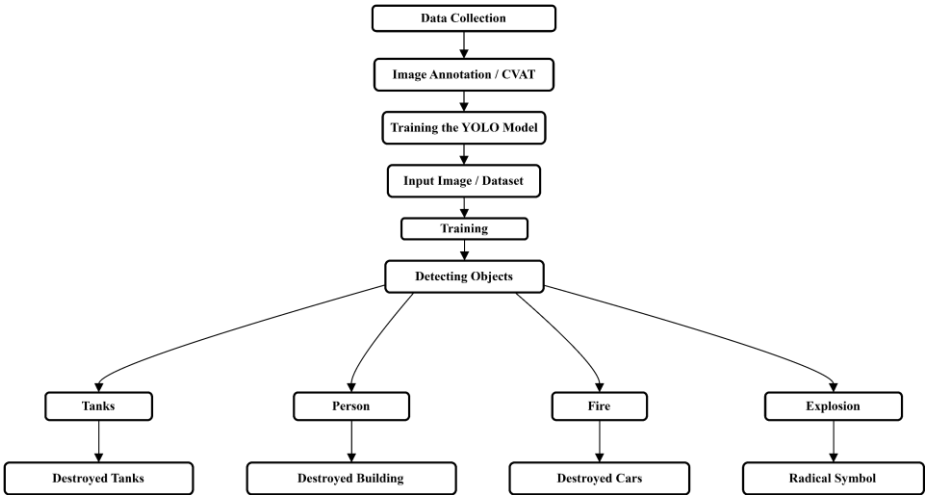

Figure 3: Workflow of Methodology

In the implementation phase of this research, we compared systematically the performance of three different versions of YOLOv8, namely YOLOv8s, YOLOv8m and YOLOv8x on a custom-labeled cyber extremism database. Finally, the Ultralytics YOLOv8 framework was

used to train the models since it is designed to be simple with great deployment capabilities and support for anchor-free detection. The training setup was based on Google Colab with NVIDIA A100 GPU and 53GB RAM, to provide high-speed computation and memory stability. All the models were trained under the same circumstances: 100 epochs, batch size 16, and input resolution of 640×640 pixels. These contributions were incorporated within a pipeline that used key features. To focus on refiner the precision, these augmentations are disabled for the last 10 epochs, which were applied in the last 10 epochs and merged them in initial epochs. The models were trained usingMomentum Stochastic Gradient descent and CIoU loss for bounding box regression and bianry cross-entropy loss for class predictions up to 100 epochs. The models were validated on a holdout validation dataset using standard metrics (Precision, Recall, mAP@50, mAP@50–95) after training. The outputs were visualized using bounding boxes and confidence scores for qualitative analysis. Trained models were exported in. pt (torch best practice) for YOLO inference and onnx (open Network exchange) for cross-platform compatibility. This enabled real-time performance testing, with inference speed per image also measured. YOLOv8s offered less latency with a rapid prediction time while YOLOv8x was the highest in detection accuracy, especially for small and overlapping objects, indicating the necessity of each model in the context of its application. Such a structured and consistent implementation across the different data sources used in this research facilitated a fair comparison that is only made possible through the reliability of the consistency of the data collection procedure to allow for valid conclusions to be made toward the practical applications of YOLOv8 in cyber extremism content detection.

An overview of the study includes descriptions of three YOLOv8 models—YOLOv8s, YOLOv8m, and YOLOv8x, that were trained on a custom dataset of pictures relevant to cyber extremism. All models were evaluated under the same training and hardware conditions for a fair and consistent performance comparison. The results were evaluated using standard object detection measures: Precision, Recall, mAP@50, and mAP@50–95.

Amongst all models, the YOLOv8x produced the highest object detection accuracy with mAP@50: 0.987 and mAP@50–95: 0.893, highlighting its robustness in detecting the various types of objects, which include small and partially occluded objects as well as images of poor quality. Its precision (0.977) and recall (0.971) scores suggested high reliability in both accurate object identification and avoiding false positive errors. On the other hand, YOLOv8x's inference time took the longest time to process with 17.3 milliseconds per image, appropriate for offline batch processing or forensic review but not ideal for real-time application.

The YOLOv8s model achieved higher speed with the least computational effort and an inference time suitable for real-time applications. It racked up a mAP@50 of 0.985 which is lower than YOLOv8x, but acceptable for practical application. Its lightweight architecture made it practical to deploy surveillance systems or automated content moderation tools.

The YOLOv8m model provided a compromise between the two, delivering good accuracy while consuming relatively few resources. It did very well in most categories but also didn't outshine YOLOv8x in precision or YOLOv8s in speed. It is a false positive suppression technique where both detection quality and latency are efficient.

In class-specific analysis, all models excelled in crucial categories such as Tank, Explosion, and Military Vehicle, with YOLOv8x attaining a mAP@50–95 score of 0.995 in these categories. YOLOv8: Real-time Object Detection with High Accuracy Across Various Object Types

The results validate the scalability of YOLOv8 models for cyber extremism object recognition. YOLOv8x is best suited for scenarios which require maximal accuracy, such as evidence review, or policy enforcement, while YOLOv8s is better for systems requiring fast, responsive detection, such as live monitoring or filtering. The selection between them is linked by the hardware used, timeline, and desired accuracy.

And the results also underline the significance of clean annotations and a well-defined training process. Still, with a smaller dataset than a lot of other challenges, the models still had great performance with more consistent labeling, realistic images with different variations, and performing augmentation during training. These further cements the idea that model performance lies as much in the architecture as in the data fed into it.

Future experiments need to employ larger and more diverse datasets to evaluate scalability and generalization. In addition, adversarial robustness should be examined, and multimodal detection using NLP methods can be fused to detect extremist content in visual and textual forms.

The YOLOv8x model performed with the following metrics:Mean Average Precision (mAP50): The model achieved a score of mAP 0.987, meaning that it detected and classified 98.7% of objects correctly with 50% Intersection over Union (IoU).

Precision & Recall: The model showed high precision and recall across multiple classes indicating the model's robustness and ability to detect and identify the object accurately with minimal false positive and negative detections.

The results in figure 4 show the performance results of the YOLOv8 models, YOLOv8s, YOLOv8x, and YOLOv8m, each with their unique dataset accuracy and inference speed. Overall precision (P) and recall (R) for YOLOv8x were 0.977 and 0.971, respectively, with a mean average precision at 50% IoU (mAP@50) of 0.987 and a mean average precision (mAP@50-95) at IoU from 50% to 95% of 0.893. Furthermore, this model required a total inference time of approximately 17.3 milliseconds per image, which is perfect for real-time applications as it can quickly process images. All models were trained with 100 no of epochs the precise, recall, mAP are shown in Table 2, 3, and 4.

Table 2: YOLOv8s Performance

No. of Epoch	Class Loss	Precision	Recall	mAP50	mAP50-95
0	5.1216	0.38219	0.16246	0.10234	0.05248
25	1.1009	0.81459	0.63312	0.77023	0.47513
50	0.70843	0.93627	0.89301	0.95634	0.73326
75	0.70485	0.93626	0.95807	0.98353	0.84928
99	0.50022	0.96918	0.97379	0.98408	0.88808

Table 3: YOLOv8m Performance

No. of Epoch	Class Loss	Precision	Recall	mAP50	mAP50-95
0	3.999	0.74813	0.15606	0.1388	0.0714
25	1.4158	0.73838	0.59362	0.66039	0.39994
50	1.0521	0.78728	0.83131	0.87141	0.62905
75	0.74136	0.9251	0.90964	0.96813	0.7824
99	0.40985	0.95969	0.95129	0.9813	0.85061

Table 4: YOLOv8x Performance

No. of Epoch	Class Loss	Precision	Recall	mAP50	mAP50-95
0	3.6808	0.71761	0.18686	0.1425	0.077
25	1.4734	0.74742	0.64871	0.68964	0.42559
50	0.951	0.90037	0.8206	0.91781	0.69129
75	0.6875	0.94136	0.93978	0.98033	0.82655
99	0.33538	0.97655	0.97093	0.98692	0.89265

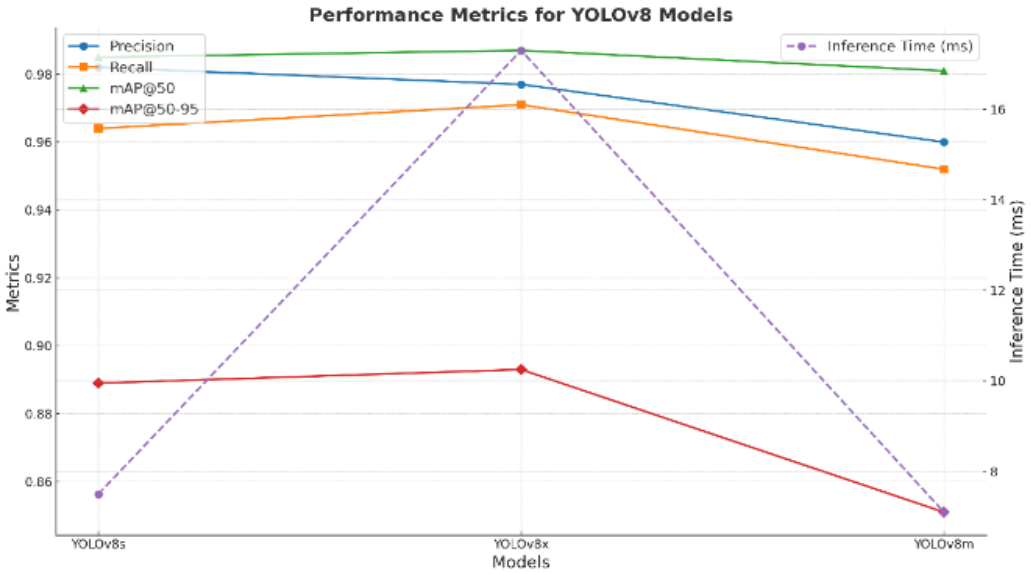

Figure 4: Performance Metric for YOLOv8 Models

Class-specific performance analyses revealed that all models achieved high precision and recall for critical categories such as "Tank," "Destroyed car," and "Explosion." However, YOLOv8s and YOLOv8x particularly excelled in maintaining consistent accuracy across most classes, with YOLOv8x displaying superior metrics for "Tank" (mAP@50-95 of 0.995) and "Military car" (mAP@50-95 of 0.995). Figure 5 shows object detections.

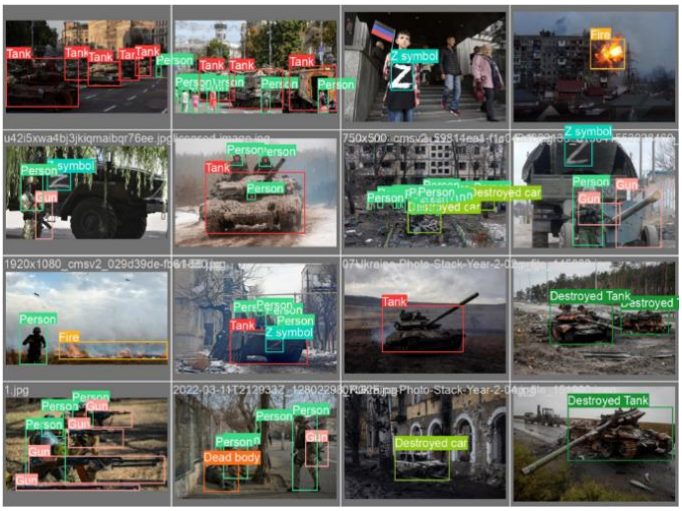

Figure 5: Object detection

As mentioned in table 2 for comparison of the three models, YOLOv8s achieved 0.985 mAP when trained for 100 epochs. The model can detect different objects; the model we trained for epoch 100. The loss of the model should be as low as possible. We can also deduce the way the model's precision, recall, and mAP change regarding accuracy. A graph in figure 6 shows performance of YOLOv8s for 100 epochs.

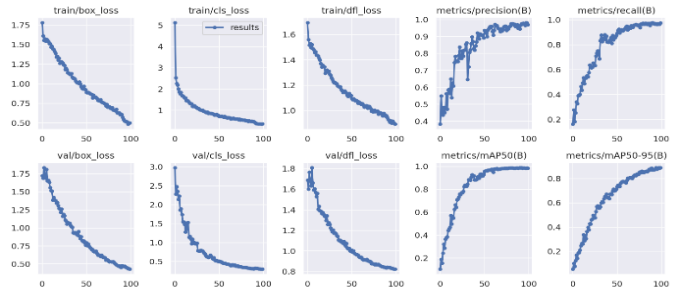

Figure 6: Results of YOLOv8s for 100 epochs.

Here, we explore the performance of three different versions (YOLOv8s, YOLOv8m, YOLOv8x) of the YOLOv8 model in detecting cyber extremism images using a custom dataset. This ensured that models were trained and evaluated in the same conditions, focusing on important metrics like precision, recall, and mean Average Precision, mAP. The results showed that the three models used for detection and classification of objects related to cyber extremism had high accuracy, particularly the mAP50 score of 0.987 of the YOLOv8x model confirming exceptional detection accuracy. However, the YOLOv8s achieved a perfect balance among the models tested, rendering the best performance in terms of detection and inference speed, making them very suitable for real-time applications [17].

Detecting cyber extremism content from visual data is becoming an increasingly challenging task, and the results of the study suggest that YOLOv8 models may help in furthering efforts to address this growing challenge. This ability to quickly and accurately analyze large volumes of data makes these models useful in determining and preventing extremist content online for law enforcement and tech companies. The YOLOv8x detector, with its exceptional accuracy, is well-suited for more performance-critical systems, such as forensic analysis or in-depth content moderation. In contrast, due to its rapid inferences, YOLOv8s can be deployed in real-time applications like live feed surveillance or immediate content moderation in social media systems.

While the results were encouraging, some challenges and limitations were noticed in the study. Data: One of the biggest challenges is around data, both quality and accessibility. The small custom dataset used in this study is effective, but it only represents a small part of the overall content regarding cyber extremism on the internet. More work needs to be done on diversifying the data sets to include a larger variety of objects, more complex scenarios, and broader ranges of image quality. Moreover, adopting AI technology for surveillance and content regulation raises ethical speculation. To ensure responsible deployment, issues such as algorithmic bias, privacy concerns, and potential use-case misuse need to be overcome.

Future work might involve making the models more robust to adversarial attacks. With cyber extremism, there are always new techniques since images and videos can be manipulated, so it is critical to create models that are resilient to these types of issues. Methods like adversarial training, data augmentation and adding extra elements of security may increase the robustness of the models.

The importance of interdisciplinary collaboration in combating cyber extremism is among the key takeaways from the study. Holistic and effective solutions can emerge when we couple skills from computer science, cybersecurity, social sciences, ethics, and more. For example, applying NLP methods along with image processing would be able to detect extremist content across different modalities thereby using a more holistic approach for content moderation.

YOLOv8x is the most accurate, while YOLOv8s is the most practical in a real-time object detection scenario. We believe the results and findings presented in this study will add to the research on deep learning in fighting cyber extremism and will serve as a steppingstone for future studies in this critical area. We can make these models even better by taking care of data quality, ethical issues and robustness of models, which will ultimately contribute to empowering our future and a safer digital world. Additionally, future research should investigate the integration of these models into larger cybersecurity architectures, allowing for more proactive and holistic responses to the continuously changing menace of cyber extremism.

Acknowledgement

This work was supported by the IQ Group Holding.

References

[1] Winter C, Neumann P, Meleagrou-Hitchens A, Ranstorp M, Vidino L, & Fürst J. Online Extremism: Research Trends in Internet Activism, Radicalization, and Counter-Strategies. International Journal of Conflict and Violence (IJCV). 2020;14:1-20 https://doi.org/10.4119/ijcv-3809

[2] Montasari R. The Impact of Technology on Radicalisation to Violent Extremism and Terrorism in the Contemporary Security Landscape. Advanced Sciences and Technologies for Security Applications: Springer International Publishing; 2024. p. 109-133. http://dx.doi.org/10.1007/978-3-031-50454-9_7

[3] M M, Rajender U, A T, & S Rumale A. Real-time object detection in videos using deep learning models. Ictact Journal on Image and Video Processing. 2023;14(2):3103-3109. Doi: https://doi.org/10.21917/ijivp.2023.0441

[4] Simmons A, & Vasa R. Garbage in, garbage out: Zero-shot detection of crime using large language models. arXiv preprint arXiv:2307.06844. 2023 https://doi.org/10.48550/arXiv.2307.06844

[5] Berhoum A, Meftah MCE, Laouid A, & Hammoudeh M. An Intelligent Approach Based on Cleaning up of Inutile Contents for Extremism Detection and Classification in Social Networks. ACM Transactions on Asian and Low-Resource Language Information Processing. 2023;22(5):1-20. Doi: https://doi.org/10.1145/3575802

[6] Aldera S, Emam A, Al-Qurishi M, Alrubaian M, & Alothaim A. Online Extremism Detection in Textual Content: A Systematic Literature Review. IEEE Access. 2021;9:42384-42396. Doi: https://doi.org/10.1109/access.2021.3064178

[7] Whang SE, Roh Y, Song H, & Lee J-G. Data collection and quality challenges in deep learning: a data-centric AI perspective. The VLDB Journal. 2023;32(4):791-813. Doi: https://doi.org/10.1007/s00778-022-00775-9

[8] Saheb T. "Ethically contentious aspects of artificial intelligence surveillance: a social science perspective". AI Ethics. 2023;3(2):369-379. Doi: https://doi.org/10.1007/s43681-022-00196-y

[9] Nguyen HH, Kuribayashi M, Yamagishi J, & Echizen I. Detecting and Correcting Adversarial Images Using Image Processing Operations. arXiv preprint arXiv:1912.05391. 2019 https://doi.org/10.48550/arXiv.1912.05391

[10] Yadav DK, Mundra K, Modpur R, Chattopadhyay A, & Kar IN. Efficient detection of adversarial images. arXiv preprint arXiv:2007.04564. 2020 https://doi.org/10.48550/arXiv.2007.04564

[11] Torregrosa J, Bello-Orgaz G, Martinez-Camara E, Del Ser J, & Camacho D. A survey on extremism analysis using natural language processing. arXiv preprint arXiv:2104.04069. 2021 https://doi.org/10.48550/arXiv.2104.04069

[12] Choudhary K. Novel Approach to Image Steganalysis (A Step against Cyber Terrorism). IOSR Journal of Computer Engineering. 2012;2(5):16-28. Doi: https://doi.org/10.9790/0661-0251628

[13] Cohen K, Johansson F, Kaati L, & Mork JC. Detecting Linguistic Markers for Radical Violence in Social Media. Terrorism and Political Violence. 2013;26(1):246-256. Doi: https://doi.org/10.1080/09546553.2014.849948

[14] VidalMata RG, Saboia P, Moreira D, Jensen G, Schlessman J, & Scheirer WJ. On the effectiveness of image manipulation detection in the age of social media. arXiv preprint arXiv:2304.09414. 2023 https://doi.org/10.48550/arXiv.2304.09414

[15] Han Y, Karunasekera S, & Leckie C. Image Analysis Enhanced Event Detection from Geo-Tagged Tweet Streams. Lecture Notes in Computer Science: Springer International Publishing; 2020. p. 398-410.10.1007/978-3-030-47426-3_31. http://dx.doi.org/10.1007/978-3-030-47426-3_31

[16] Jayaram D, Gopalachari MV, Rakesh S, Sai JS, & Kumar GK. Fake face image detection using feature network. International journal of health sciences. 2022:3027-3039. Doi: https://doi.org/10.53730/ijhs.v6ns5.9310

[17] Albdairi M, & Almusawi A. Examining the Influence of Autonomous Vehicle Behaviors on Travel Times and Vehicle Arrivals: A Comparative Study Across Different Simulation Durations on the Kirkuk-Sulaymaniyah Highway. International Journal of Automotive Science And Technology. 2024;8(3):341-353. Doi: https://doi.org/10.30939/ijastech..1480916

[18] Kumar V, Sharma S, Kumar C, & Sahu AK. Latest Trends in Deep Learning Techniques for Image Steganography. International Journal of Digital Crime and Forensics. 2023;15(1):1-14. Doi: https://doi.org/10.4018/ijdcf.318666

[19] Tamang S, Sen B, Pradhan A, Sharma K, & Singh VK. Enhancing COVID-19 Safety: Exploring YOLOv8 Object Detection for Accurate Face Mask Classification. International Journal of Intelligent Systems and Applications in Engineering. 2023;11(2):892 - 897 https://ijisae.org/index.php/IJISAE/article/view/2966

[20] Prusty MR, Tripathi V, & Dubey A. A novel data augmentation approach for mask detection using deep transfer learning. Intell Based Med. 2021;5:100037-100037. Doi: https://doi.org/10.1016/j.ibmed.2021.100037

[21] Han Z, Huang H, Fan Q, Li Y, Li Y, & Chen X. SMD-YOLO: An efficient and lightweight detection method for mask wearing status during the COVID-19 pandemic. Comput Methods Programs Biomed. 2022;221:106888-106888. Doi: https://doi.org/10.1016/j.cmpb.2022.106888

[22] Qin B, & Li D. Identifying Facemask-Wearing Condition Using Image Super-Resolution with Classification Network to Prevent COVID-19. Sensors (Basel). 2020;20(18):5236. Doi: https://doi.org/10.3390/s20185236

[23] Jignesh Chowdary G, Punn NS, Sonbhadra SK, & Agarwal S. Face Mask Detection Using Transfer Learning of InceptionV3. Lecture Notes in Computer Science: Springer International Publishing; 2020. p. 81-90.10.1007/978-3-030-66665-1_6. http://dx.doi.org/10.1007/978-3-030-66665-1_6

[24] Loey M, Manogaran G, Taha MHN, & Khalifa NEM. Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. Sustain Cities Soc. 2021;65:102600-102600. Doi: https://doi.org/10.1016/j.scs.2020.102600

[25] Taneja S, Nayyar A, Vividha, & Nagrath P. Face Mask Detection Using Deep Learning During COVID-19. Lecture Notes in Networks and Systems: Springer Singapore; 2021. p. 39-51.10.1007/978-981-16-0733-2_3. http://dx.doi.org/10.1007/978-981-16-0733-2_3

[26] Harriat Christa G, J J, K A, & Sagayam KM. CNN-based Mask Detection System Using OpenCV and MobileNetV2. 2021 3rd International Conference on Signal Processing and Communication (ICPSC); 2021/05/13: IEEE; 2021. p. 115-119.10.1109/icspc51351.2021.9451688. http://dx.doi.org/10.1109/icspc51351.2021.9451688

[27] He K, Zhang X, Ren S, & Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016/06: IEEE; 2016.10.1109/cvpr.2016.90. http://dx.doi.org/10.1109/cvpr.2016.90

[28] Kumar A, Kalia A, Verma K, Sharma A, & Kaushal M. Scaling up face masks detection with YOLO on a novel dataset. Optik. 2021;239:166744. Doi: https://doi.org/10.1016/j.ijleo.2021.166744

[29] Degadwala S, Vyas D, Chakraborty U, Dider AR, & Biswas H. Yolo-v4 Deep Learning Model for Medical Face Mask Detection. 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS); 2021/03/25: IEEE; 2021. p. 209-213.10.1109/icais50930.2021.9395857. http://dx.doi.org/10.1109/icais50930.2021.9395857

[30] Abbasi S, Abdi H, & Ahmadi A. A Face-Mask Detection Approach based on YOLO Applied for a New Collected Dataset. 2021 26th International Computer Conference, Computer Society of Iran (CSICC); 2021/03/03: IEEE; 2021. p. 1-6.10.1109/csicc52343.2021.9420599. http://dx.doi.org/10.1109/csicc52343.2021.9420599

[31] Shen L, Lang B, & Song Z. DS-YOLOv8-Based Object Detection Method for Remote Sensing Images. IEEE Access. 2023;11:125122-125137. Doi: https://doi.org/10.1109/access.2023.3330844

[32] Sapkota R, Ahmed D, & Karkee M. Comparing YOLOv8 and Mask R-CNN for instance segmentation in complex orchard environments. Artificial Intelligence in Agriculture. 2024;13:84-99. Doi: https://doi.org/10.1016/j.aiia.2024.07.001

[33] Talib M, Al-Noori AHY, & Suad J. YOLOv8-CAB: Improved YOLOv8 for Real-time object detection. Karbala International Journal of Modern Science. 2024;10(1). Doi: https://doi.org/10.33640/2405-609x.3339