Crime Scene Analysis for News Headline Generation

Dr. Sandeep Varma N
Department of Machine Learning
B.M.S. College Of Engineering
Bangalore, India
sandeep.mel@bmsce.ac.in

Akash Venugopal
Department of Machine Learning
B.M.S. College of Engineering
Bangalore, India
akashvg2001@gmail.com

Ayush Sharma
Department of Machine Learning
B.M.S. College of Engineering
Bangalore, India
ayush.sh314@gmail.com

***Abstract*— The news headline generation of crime is a tedious task in journalism reporting that requires analysis and creative skills. Often, human observation and reasoning is subject to error or fallacies, and it is a time-consuming task. Hence an automated headline generation mechanism would greatly benefit media teams. The traditional approach of news headline generation primarily relies on textual data and not on images, which may not always provide a comprehensive understanding of the story. Generating headlines from images requires a deep understanding of the visual content and the ability to extract relevant information. Thus, there is a need for a system that can leverage the power of image detection to automatically generate news headlines that accurately reflect the visual content of an image. This research, which has been proposed by this paper could potentially revolutionize the way news headlines and articles are generated, as it would allow for a more nuanced and accurate representation of the crime scene.**

Keywords— News headline generation, News Article generation, image extraction, YOLOv5, GPT2, text generation

I. Introduction

Utilizing the concept of Object Detection for accurate recording of crime scenes is proving to be immensely useful in Digital Forensics. The analysis of visual evidence obtained from a crime scene can be done with little to no human intervention. With the deployment of real-time object detection models like the YOLO[1], one can identify objects and document the observations of a crime scene. The increasing popularity of NLP algorithms for the purpose of text generation has also made its way into the law enforcement taskforce, where the combined deployment of these two models helps in generating precise and reliable news headlines of the crime scene.

News headline generation for crime is a tedious task in journalism reporting that requires analysis and creative skills. Often, human observation and reasoning is subject to error or fallacies, and it is a time-consuming task. Hence this work has tried to devise an automated news headline generation mechanism which can be followed by media teams to obtain coverage on small scale and bigger crimes happening in the neighborhood. The domain of "news headline generation from crime scene images" is a specific application of computer vision, natural language processing, and journalism.

This task involves developing a system that can automatically generate news headlines from crime scene images. The system must be able to analyze and interpret images to identify key details, such as the location, type of crime, and any significant visual elements. It must also be able to generate a concise and attention-grabbing headline that accurately represents the crime scene. Additionally, the system may need to incorporate additional information, such as eyewitness accounts or official statements, to provide context and accuracy to the headline. To develop such a system, expertise in computer vision, natural language processing, journalism, and machine learning may be required. The resulting system could provide an efficient and timely way to report on breaking news stories and improve the speed and accuracy of news reporting.

The main objective is to develop an automated system that can quickly and accurately generate news headlines based on crime scene images. The system should be able to analyze and interpret images to identify relevant details, and generate a headline that accurately represents the crime scene. The system should be designed to be fast and efficient, enabling news organizations to quickly report on breaking news stories. Additionally, the system should prioritize accuracy and provide context to the headline by incorporating additional information, such as eyewitness accounts or official statements. By automating the headline generation process, news organizations can improve the speed and accuracy of their reporting, enabling them to provide timely and relevant information to their audiences.

II. Related Work

Recent work in the field of Crime Scene analysis has been carried out with a significant role of Object detection. Various deep neural network architectures have been proposed for the detection of objects in an image of a crime scene. Deep neural network architectures such as GoogLeNet[2], AlexNet[3], VGG - Net[4] have been trained to classify objects in an image with satisfactory results. Convolutional Neural Networks are extensively used for processing images and for classification of objects. A typical CNN architecture consists of a Convolution layer, Pooling layer, and Fully Connected Layer. Convolution Layers are followed by activation function units, generally ReLu(Rectified Linear Unit). CNNs preserve spatial information in images which is a very important feature for the task of Crime scene analysis.

Surajit et al. [5] developed a model with pre-trained VGG-16 architecture of Faster R-CNN (Region-Based Convolutional Neural Network) which was trained on MS-COCO dataset. Alayesanmi et al.[6] developed an object detection model based on YOLO to perform detection of objects at a crime scene without human involvement or any external control. Drishti et al.[7] worked on detection of crime scenes by detection of three classes of crimes, YOLO was used for the same. D.S. Devishree et al.[8] developed a

Deep Neural Network to identify three classes of objects, namely gun, knife, and blood. ReLu was used for denoising. A method for generating sentences from "keywords" or "headwords" is devised by Kiyotaka et al. [9]. This method consists of two main parts, candidate-text construction and evaluation.

III. METHODOLOGY

This work uses a YOLOv5[10] model for Object Detection whose standard code can be implemented using Python in a local text editor. Further, Python's library KeytoText[11] is used for sentence generation. The final text generation model used is GPT-2[12], which is a general-purpose NLG learner. The sophisticated underlying architectures of YOLO and Transformers assist us in implementing these pre-built models. Detailed methodology is shown in figure 1.

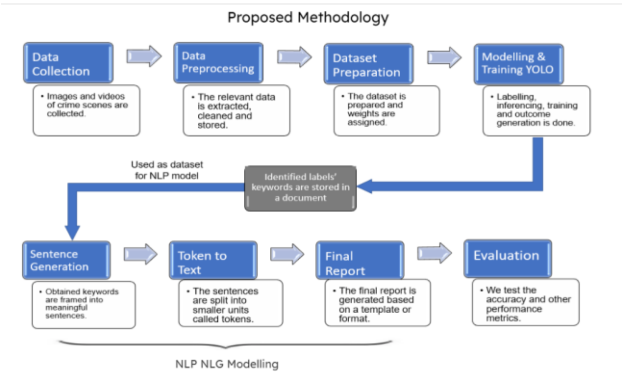

Figure 1 : Proposed Methodology

A. Implementation

The implementation of our work has been carried out on VS Code and Google Colab using Python. GPU instances of Google Colab were used for training the object detection model.

The implementation of this project consists of 3 phases:

- Phase 1: Implementation of YOLO v5 model for Object Detection.
- Phase 2: Extraction of labelled image tags and Sentence Generation using KeytoText library.
- Phase 3: Paragraph Generation using GPT2.

The steps involved as shown in Figure 2 and Figure 3 in the each of these phases are elucidated below:

Phase 1: Implementation of YOLO v5 model for Object Detection

YOLO v5 is a prebuilt object detection model that is loaded from the official git repository. The torch library is used.

For each detection from each output layer, the confidence, class id, bounding box parameters are obtained and the weak detections are ignored (with confidence < 0.3).
The output of this phase will be an image with text-labelled objects that are called "tags."

Phase 2: Extraction of tags and Sentence Generat on using KeytoText library

The tags displayed for the image obtained in Phase 2 are extracted into a Data frame, and from there, they are made into a list which is further fed to the KeytoText library to obtain the sentence. KeytoText is based on the T5[13] Model, which is a Transformer based architecture that uses a text-to-text approach. Every task – including translation, question answering, and classification – is cast as feeding the model text as input and training it to generate some target text.

Phase 3: Paragraph Generation using GPT-2:

Generative Pre-trained Transformer 2 (GPT-2) is an open-source artificial intelligence model created by Open AI. GPT-2 translates text, question answering, summarizes passages, and generates text output. The sentence generated in Phase 2 is fed to the GPT-2 model which generates a paraphrased headline, the length of which is up to 3-4 lines long.

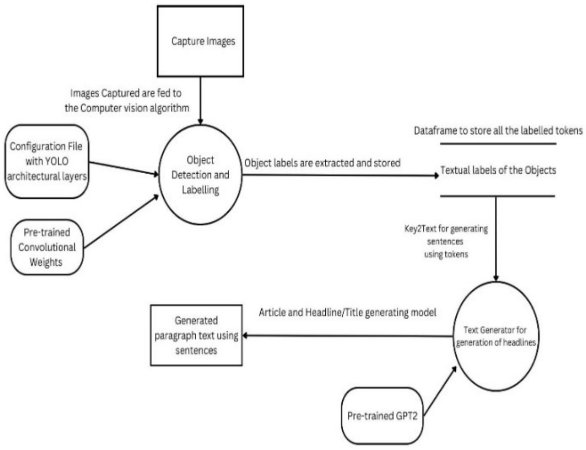

Figure 2 : Data flow diagram

HuggingFace's transformers library is used, and all the dependencies are stored in transformers.
GPT2tokenizer is used to tokenize the words to be used in paragraph generation and the encode method of tokenizer is used to encode the sentence into a vector.

- Max length of output paragraph-1000 words.
- Based on- Beams Search algorithm.
- n-gram size- 4 words are fed as tokens to generate a relevant paragraph.

Decoding of the output vector takes place. and the final output paragraph is generated by the model.

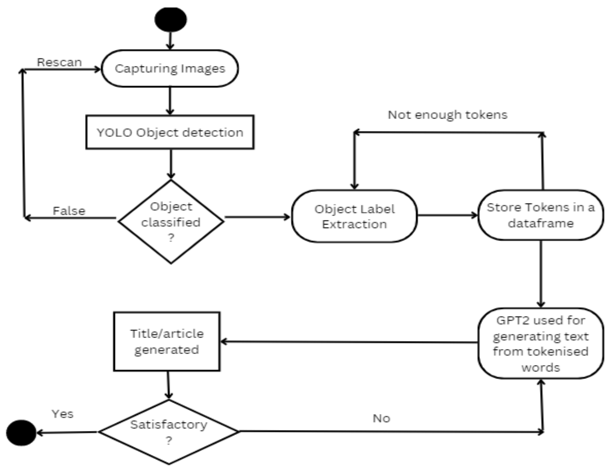

Figure 3 : Control Flow Diagram

B. ALGORITHM :

Algorithm: Crime Scene Analysis and Text Generation

Input: Image captured from camera

Output: News headline generated for the observed crime scene

Assumptions:
- Image is captured clearly for the algorithm to analyze and generate text.
- Image crowding and other image clarity parameters are in place.

Begin:

If image is captured properly:
1. Perform object detection on the image to identify potential crime evidences and label the objects on it.
2. Extract labels from the objects which are labelled.
3. Store the extracted image label keywords in a Pandas dataframe.
4. Feed the tuples of the dataframe to the key2text library sentence generator.
5. Feed the generated sentence to the GPT2 model for paragraph generation which is meaningful.

End

C. TESTING AND VALIDATION

The following metrics to evaluate the model:
- Precision - The quality of positive predictions made by the model.
- Recall- The recall measures the model's ability to detect positive samples (Higher the recall, the more positive samples detected.)
- mAP (Mean Average Precision) - metric used to evaluate object detection models such as Fast R-CNN, YOLO, Mask R-CNN, etc. The mAP compares the ground-truth bounding box to the detected box and returns a score. The higher the score, the more accurate the model is in its detections.

Class	Images	Instances	P	R	mAP50	mAP50-95
All	121	441	0.822	0.775	0.842	0.459
Gun	121	64	0.939	0.968	0.987	0.648
Bullet shell	121	72	0.767	0.549	0.672	0.289
Blood	121	272	0.693	0.672	0.732	0.339
Knife	121	33	0.891	0.909	0.977	0.561

Table 1: mAP values for different classes in the dataset

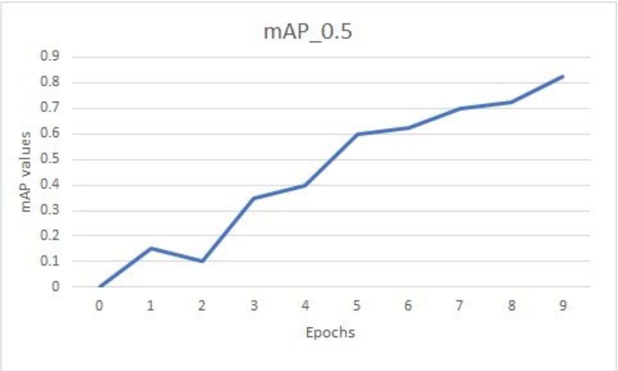

Figure 4 : Mean Absolution Precision [0.5]

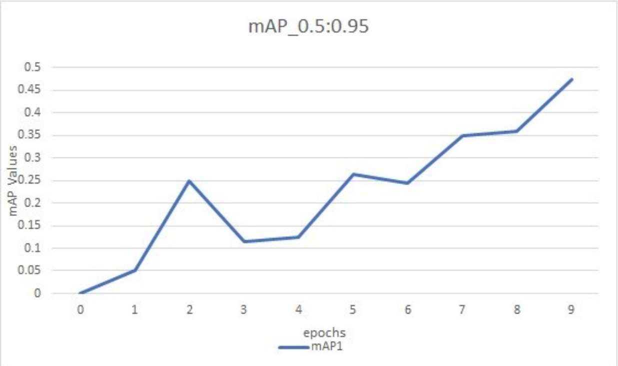

Figure 5 : mAP[0.5:0.95]

The above graphs give a brief understanding of the mean average precision obtained after the training has been carried out. Figure 4 gives a mAP value for a total of 10 epochs duration and the mAP value lands at 0.84 approximately. The x-axis gives the range of epochs [1-10].

Figure 5 provides the mAP value for each epoch averaged over IoU threshold range [0.5 : 0.05 : 0.95].

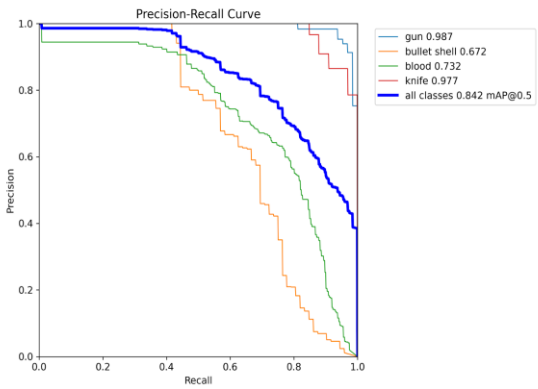

Figure 6: PR curve

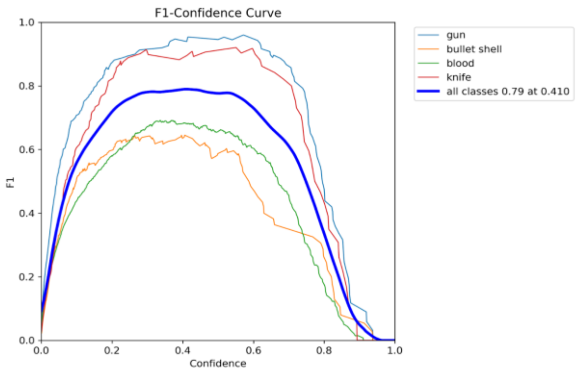

Figure 7: F1 curve

From the above, Figure 6 gives the Precision-Recall curve which measures the Precision and Recall on one scale for better analysis. It is observed that for the trained classes, the mean-averaged precision values are satisfactory and the area under the curve for each of the classes is large, which indicates that the classifier has been trained well on the dataset of crime scene images.

The darkened line in figure 6 gives the results of all the classes which were included in training and it is observed that the mAP value obtained is 0.84.

Figure 7 represents the F1-curve which gives a visualization taking into consideration of precision and recall values.
For different confidence values, F1 scores are observed for the different classes. It is observed that as confidence increases, precision goes up-to a certain optimized value after which the recall decreases which in turn makes the curve to tend downwards for higher confidence values.

IV. RESULTS AND DISCUSSION

The Phase 1 of our implementation, which is Object Detection, produces the following result:

Figure 8: Object detection result

Figure 9: Object detection result

Figure 10 : Object detection result

Figure 11: Object detection result

Images as shown in figure 8, figure 9, figure 10, figure 11 are fed to the model and the labels are generated by our proposed model. These labels are extracted from the image and stored in a panda's data frame. The stored keywords are then fed to the text generation model which resulted into textual description as shown in the table 2.

Input Sentence	**Generated Output Paragraph**
"Girl hit on the road"	"A girl was hit on the road and taken to the local hospital where she was treated for minor injuries"
"A man dead knife kitchen"	"A man was stabbed to death in a kitchen knife attack in the city's north west"
"A man found with a gun on the floor"	"A man found with a gun on the floor was pronounced dead when brought to the hospital, the police are investigating the shooting"

Table 2: GPT-2 result

It is observed that the test data, which is a set of images depicting various crime scenes, when fed to the YOLO algorithm detects certain objects of confidence>0.3. It can be deduced that these objects are detected accurately by simply viewing the images and the items present in them, i.e., a man, woman, gun, blood etc. The tags (labels) of the objects are extracted successfully and used as the "keywords" for framing a meaningful sentence. This sentence is then fed to the GPT-2 algorithm which encodes the sentence into a vector and decodes it to ultimately give an NLG output which is a relevant paragraph.

V. CONCLUSION AND FURTHER ENHANCEMENTS

This work achieves the successful detection of relevant objects present in a crime scene which may be missed out by humans. An AI generated crime scene report/headline can be a valuable addition to the work of the forensics team and cops and to the media team.

Automation of this task would allow the media personnel and editors to direct their time and efforts toward covering more stories and make their task easier. It could also help to identify the most important elements of a story and generate a headline that accurately reflects the story. This would help to improve the accuracy and speed of headline generation, and ensure that headlines are more representative of the story that they are accompanying.

- Work can be made to create bigger news articles with specific tone and sentiments -> Can be achieved by fine tuning the GPT2 model, datasets like Emolex. Object and environment analysis can be made more detailed to extract more information about the scene

- More coherent paragraphs can be made, and objects can be detected by using newer and powerful models such as GPT3 or YOLOv7. KeyToText model fails to create meaningful sentences when the frequency of tags extracted from YOLO v5 are high, this could be overcome by using better n-gram models.

REFERENCES

[1] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). Youonlu look once: Iinified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

[2] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A. (2015). Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9.

[3] Krizhevsky, A., Sutskever, I., Hinton, G.E.(2012): Imagenet classification with deep con- volutional neural networks. In: Advances in Neural Information Processing Sys- tems, pp. 1097–1105

[4] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[5] Saikia, S., Fidalgo, E., Alegre, E., & Fernández-Robles, L. (2017). Object detection for crime scene evidence analysis using deep learning. In *Image Analysis and Processing-ICIAP 2017: 19th International Conference, Catania, Italy, September 11-15, 2017, Proceedings, Part II 19* (pp. 14-24). Springer International Publishing.

[6] Alayesanmi femi Samson, Francisca Oladipo & Emeka Ogbuju. (2018) .An object detection and classification model for Crime Evidence Analysis using YOLO. In Proceedings on Big Data Analytics & Innovation Volume 3, pp. 18-26

[7] Jalgaonkar, D., Gund, J., Patil, N., & Phadke, M. (2020). Detecting Crime Scenes using ML. *International Research Journal of Engineering and Technology(IRJET)*, *Volume:07 Issue:05(2020)*

[8] Devishree D. S., Divakar K. M., Ashini K. A., Arnav Singh Bhardwaj, Sheikh Mohammad Younis (2019). Crime scene prediction and analysing its accuracy with frames using deep neural network. *International Journal of Advance Research, Ideas and Innovations in Technology*, 5(2) www.IJARIIT.com.

[9] Uchimoto, K., Sekine, S., & Isahara, H. (2002). Text generation from keywords. In *COLING 2002: The 19th International Conference on Computational Linguistics*.

[10] https://github.com/ultralytics/yolov5

[11] https://pypi.org/project/keytotext/

[12] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI* blog, 1(8), 9.

[13] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., … & Liu, P. J. (2020). Exploring the limits of transfer learning with unified text-to-text transformer. *The Journal of Machine Learning* Research, 21(1), 5485-5551.