Open-source Image Analysis Tool for Object Identification and Distance Estimation for Crime Scene Investigation

Patrick Mark Andrei M. Castro
Department of Computer Science
Adamson University
Manila, Philippines
patrick.mark.andrei.castro@adamson.edu.ph

Angelo D. Erasmo
Department of Computer Science
Adamson University
Manila, Philippines
angelo.erasmo@adamson.edu.ph

Julius Fredrick Nocon
Department of Computer Science
Adamson University
Manila, Philippines
julius.fredrick.nocon@adamson.edu.ph

Jhon Henry S. Bayquen
Department of Computer Science
Adamson University
Manila, Philippines
jhon.henry.bayquen@adamson.edu.ph

***Abstract*—This capstone project is focused on the detection, classification of objects and providing useful information about the image for crime scene investigation with the use of YOLOv4 image detection model and using pairwise Euclidean distance formula. The output is a single page web application along with its function and features such as the objects on the image, object count, and distance between the objects. The data from the image can be exported to CSV and different objects can be cropped and downloaded. The testing datasets were gathered from different sources such as MIT Indoor Scenes, and ExDARK. Test data will be used in the process of object detection and classification using YOLOv4 for Crime Scene Investigation. By testing the object detection on the given dataset, it helped the researchers gather results about the confidence, accuracy of the model and provide conclusion and recommendation. The study concludes that on dark or low light images, the object detection has a high average confidence level on the objects detected, which means that the objects found on the dark images are likely correct. The object detection model has a low accuracy on detecting small objects. The web application object detection together with its features can be helpful in crime scene investigation. The image detection provides acceptable confidence level and accuracy.**

Keywords— Object detection · Convolutional neural network · Deep learning · Image analysis · Crime scenes · Web application*

I. Introduction

Object detection is one of many pieces that classifies computer vision. It is an image processing that deals with detecting instances of objects from a particular class in an image or video. It aids in many ways as it is a general tool for specific uses.

Images is a visual representation of something which stores data or information. And in other matters, images can be a big help to gather information or ideas [1]. For example in the fields of crime investigation. Images are important to a forensic investigator. It helps the investigator to understand how, why, when, did the crime happen and other important queries to solve cases. Some crime scenes lacks evidence and can lead into an incompetent act as images cannot be determined. Images provide visual documentation of the scene and locations of the evidence within the scene. Photographs taken at the crime scene allowed the IO to recreate that scene for later analysis or for the use in the courtroom [2]. Several types of evidence are commonly found at crime scenes. Each item of evidence must be documented photographically, showing its location and appearance. Other than images, in most cases distance is measured and analyzed to produce specific data.

For this project, it can be used as an Open Source Intelligence (OSINT) tool. It will be useful in the field of Forensics, specifically in Crime Scene Examination. The application can be used by forensic scientists that seek and extract comprehensive information about the image of the crime scene, particularly in indoor crime scenes.

II. Literature Review

A. Crime Scene Investigation

The goal of a crime scene investigation is to document the conditions at the location and collect any tangible evidence which can explain what might have happened.

Crime scene investigation adheres to a set of acceptable rules and methods that ensure that all physical evidence is identified and investigated. The basic crime scene procedures are physical evidence recognition, documentation, proper collection, packaging, preservation, and, finally, scene reconstruction [3].

B. Crime Scene Investigation Tools (manual/hands-on)

As a crime scene investigator it is important to have the right tools in order to process a crime scene properly, there are many items that make up a crime scene kit such as evidence collection tools, protective gear, and other items. Other types of kit that are essential for crime scene personnel are blood collection kit, bloodstain pattern kit, excavation kit, fingerprint kit, impression kit, trajectory kit, pattern print lifter kit, trace evidence collection kit, and etc. These are needed in order for the investigator to find and gather as much evidence safely without damaging the evidence [4].

C. Image Analysis CSI Tools (computer-based tools)

In today's era, many technologies arise in the field of Forensics. Technologies such as: (1) High-Speed Ballistics Photography is used by ballistics specialists who often use high-speed cameras in order to understand how bullet holes, gunshot wounds and glass shatters are created; (2) Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA-ICP-MS) is used when broken glass is related in a crime, putting together even the smallest fragments can be crucial in determining significant information such as the direction of bullets, the force of impact, or the type of weapon used. The LA-ICP-MS machine breaks glass samples of almost any size to their atomic structure using its highly sensitive isotopic recognition capability [5].

D. Image Analysis in Crime: Progress, Problems and Prospects

(1) Detection, (2) prevention, and (3) forensics are the three most essential components of crime. Before proceeding, keep in mind that the certainty of detection will also serve as a deterrent

to criminal activity. Similarly, tactics such as surveillance have the potential to.

Detection of crime may be beneficial in anticipating crime and therefore, if appropriate steps are implemented in a timely manner, it can be eliminated. (e.g. the authorities can be alerted to people acting suspiciously) [6].

E. *Object Detection for Crime Scene Evidence Analysis Using Faster R-CNN*

In this study, they describe a real-time system based on a Faster R-CNN (Region-based Convolutional Neural Network) that automatically recognizes objects in an indoor environment. The study deployed the suggested approach to a subset of ImageNet comprising 12 object types using Karina dataset to see how effective it was. In Nvidia-TitanX GPU, the study achieved an average accuracy of 74.33 percent and a mean time to detect objects per image of 0.12 seconds [7].

F. *Distance Determination*

Distance determination testing is useful in situations where the distance a firearm was held from a target or victim is critical in determining what happened during the commission of a crime or demonstrating self-defense [8].

G. *Bullet Trajectory*

A method for determining the bullet trajectories is a technique called *Stringing*. It is used at crime scenes to determine the source of blood spatter and the trajectory of bullets.

III. Research Methodology

A. *YOLOv4*

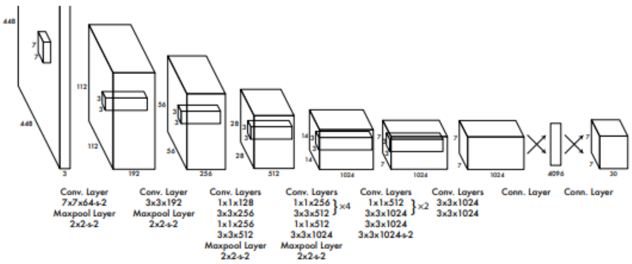


Fig. 1 YOLO Architecture

The YOLO (You Only Look Once) architecture consists of 27 CNN (Convolutional Neural Network) layers, with 24 convolutional layers, along with 2 Fully Connected layers and a final detection layer as shown in Figure 1. It splits input images into S × S grid cells and within each grid cell predicts B bounding boxes and a score for each of the C classes. Each bounding box consists of 5 predictions which are center x, center y, width, height and confidence of the bounding box. For each grid cell, there will only be one set of class scores C for all bounding boxes in that region. Thus, the output of the YOLO network will be a vector of S× S × (5B + C) numbers for each image. While the fully connected layers use the features extracted from the convolutional layers and use the information to predict the probabilities of the object and at the same time for the bounding box constructions. Then the final layer which is a YOLO detection layer is a regression that maps the output of the last fully connected layer to the final bounding box and class assignments.

The main components of YOLOv4 uses Darknet53 [9], SSP [10], PAN [11] as the neck and the for the head it uses YOLOv3 [12]. As well as other BoF (Bag of Freebies) and BoS (Bag of Specials). Bag of freebies are training methods which change the strategy for training and increase training cost. Bag of specials are methods that will increase inference cost of the model but will improve its performance [13]. Darknet P53 is an effective backbone for extracting features. It has a deep backbone with 53 convolutional layers and several advanced structures, such as: (a) Residual blocks, which add shortcut layers to make the network easier to train; (b) Inception structure, which includes 3x3, 1x1 convolutional kernels to hold the respective field while lowering the computation time. The SPP-block, and PAN path-aggregation block, are used as post-processing methods to improve accuracy. The YOLOv3 head uses a function pyramid network (FPN)-like structure. In terms of the various scales, it makes three predictions. The bounding box coordinates, confidence scores for each class, and object confidence (1 for object and 0 for non-object) are all included in the output tensor. The YOLO head's outputs are post-processed with non-maximum suppression to eliminate noise.

The object detector is based on YOLOv4 which was trained on the COCO dataset.

B. *Object Count*

The object count is the count of the different classes of objects present in the image. The total count of objects is also presented in the output.

C. *Pairwise Euclidean distance formula*

$$d(x,\ y)\ =\ \sqrt{\sum_{i=1}^{n}(y_i - x_i)^2}$$

Fig. 2 Pairwise Euclidean distance formula

The pairwise Euclidean distance formula is used for the measurement of the distances between objects. The centroids of image combinations can be used to calculate the distance of each object pair in an image.

D. *Programming Language*

The main programming language used for the application is Python. The backend of the web application was developed using Flask. And the image analysis was done in Python using OpenCV.

E. *Dataset*

The datasets used in this research study were: COCO (common objects in context) dataset [14] is a large-scale object detection, segmentation, and captioning dataset. The detector used a pre-trained model which was trained from this dataset;

MIT indoor scenes dataset [15] is a dataset containing indoor scene images, this dataset is used to test for indoor crime scenes as this dataset interconnects with the environment of the project deployment;

ExDARK dataset [16] is a dataset with images that has a low-brightness from a low-light environment. This dataset is similar to PASCAL VOC. This dataset will be useful for testing dark images in indoor scenes.

F. *Testing*

The accuracy of the image detection algorithm is tested with the MIT indoor scenes dataset and ExDARK dataset. Around 7866 images from both datasets were tested.

G. *Prototype*

The prototype of the project is a web application that contains several functions and features. Functions such as image upload, object classification, total objects count, count per object, distance between objects. And features such as object filtering, crop detected objects, and exporting data to CSV file.

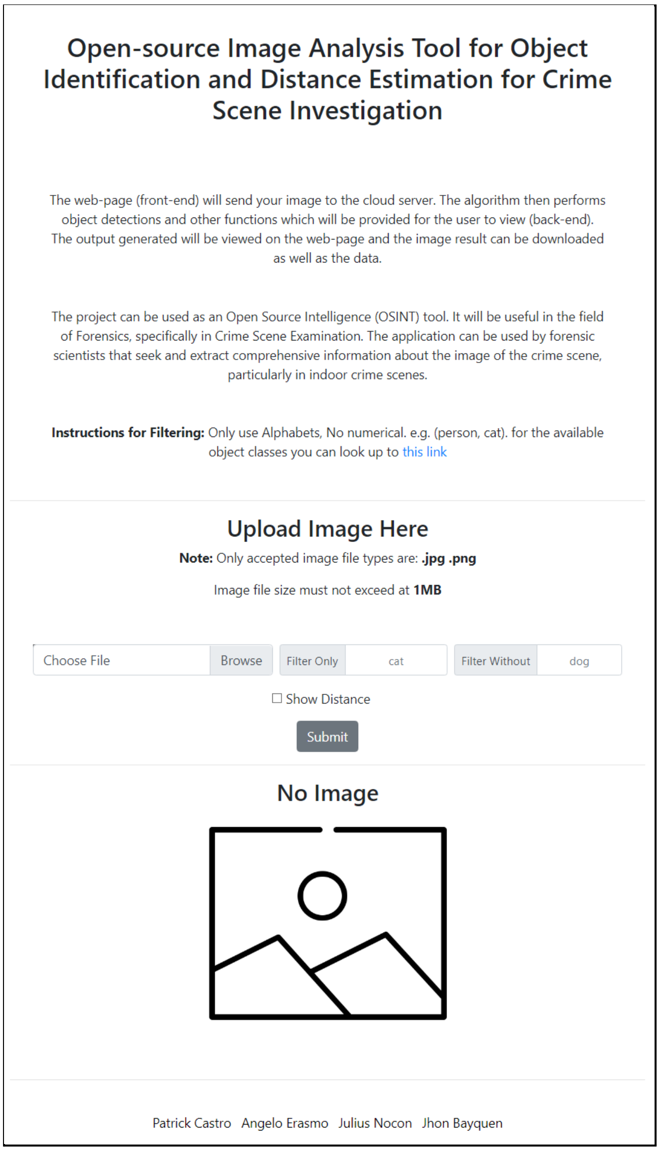


Fig. 3 Web Application

IV. Result and Discussion

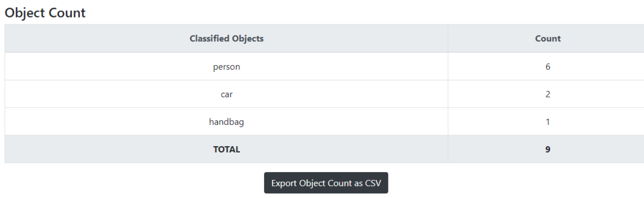
The general objective of the study is to create a tool that applies image analysis and gives information about the image and to apply distance estimation on the image that returns distance between objects, is achieved by the following:

The results are presented in two parts, the first part is the results and outputs from the developed web application, explaining the details of each function. The second part consists of testing the object detection on two dataset to determine the performance of the object detection when used in crime scene environments.

Fig. 4 Image Object Detection output

A. Object Detection

Figure 4 shows the object detection from the image detection mode, a rectangular box is shown for each object detected, as well as the class of the object and the confidence. The name of each object is unique, (eg. person1, person2) which helps to distinguish objects when doing distance estimation and exporting the data.


Object Count

Classified Objects	Count
person	6
car	2
handbag	1
TOTAL	9

Export Object Count as CSV

Fig. 5 Object count

B. Object Count

Object count (Fig. 5) shows the number of objects detected in an image, as well as the total objects in the image which can be exported as a CSV.

Fig. 6 Distance estimation

C. Distance Estimation

Distance estimation can be shown on the output depending on the user. The result will show the distance between the objects in the image, the object is highlighted with a red mark and a blue line between them with an indication of the distance shown. A table of distances between objects is also shown which can be exported as CSV.

D. Object Detection Confidence

ExDARK: The Exclusively Dark (ExDARK) dataset is a collection of 7,363 low-light images from very low-light environments to twilight (i.e 10 different conditions) with 12 object classes (similar to PASCAL VOC) annotated on both image class level and local object bounding boxes. By using this dataset, It will boost the confidence level of the model.

MIT Indoor Scenes: The database contains 67 Indoor categories, and a total of 15620 images. The number of images varies across categories, but there are at least 100 images per category. All images are in jpg format.

Class	Confidence	Detections	Total Images
Bus	91.70 %	605	527
Cup	84.35 %	661	519
Car	83.97 %	1309	638
People	81.97 %	1663	609
Bottle	80.95 %	1017	547

Bicycle	80.82 %	975	652
Chair	80.16 %	817	648
Motorbike	79.21 %	777	503
Boat	72.65 %	984	679
Table	60.78%	225	505

Table 1. Detection confidence for each object class in the ExDARK dataset.

As seen on Table 1, testing from the ExDARK dataset helped the researchers determine the confidence scores of the detection model on low-light images. The confidence score is the probability that an anchor box contains an object. The detections column is the number of detections

The table shows 10 object classes, the average confidence from each detection, the number of images and the number of detections It shows a 91.70% high confidence for Bus objects with 605 detections in 527 images; and a 60.78% low confidence for Table objects with 225 detections in 505 images.

Class	Confidence					Detections
	bathroom	bedroom	corridor	garage	kitchen	
Bicycle	-	-	62.71 %	78.54 %	-	31
Boat	-	-	-	-	-	-
Bottle	62.80 %	62.26 %	66.34 %	58.88 %	65.76 %	364
Bus	-	-	-	-	-	-
Car	-	71.89 %	-	84.03 %	-	18
Chair	77.14 %	78.90 %	73.88 %	80.56 %	81.36 %	970
Cup	66.83%	50.38 %	-	57.10 %	63.00%	143
Motorbike	-	-	-	84.86 %	-	13
People	65.79 %	68.78 %	78.72 %	88.82 %	-	82
Table	97.96 %	58.94 %	69.46 %	70.40 %	65.05 %	189
Vase	71.81 %	68.48 %	57.23 %	-	65.25 %	277
Bowl	66.29 %	59.16 %	95.91 %	51.77 %	67.84 %	251
Handbag	54.31 %	63.28 %	60.51 %	52.09 %	48.23 %	15
Hair drier	57.74 %	-	-	-	-	4
Clock	69.26 %	64.05 %	83.10 %	82.16 %	80.81 %	63
Refrigerator	96.93 %	74.07 %	60.28 %	71.43 %	82.85 %	252
Backpack	-	64.04 %	-	42.09 %	65.35 %	10
Laptop	-	77.78 %	88.29 %	-	56.92 %	15
Suitcase	-	55.79 %	56.33 %	50.36 %	-	11
Remote		70.19%	-	-	78.93 %	18
Wine glass	-	69.62%	-	-	66.92 %	40
Umbrella	-	-	40.41 %	-	-	1
Spoon	-	-	-	-	58.35 %	41
Fork	-	-	-	-	42.47 %	1
Knife	-	-	-	-	61.55 %	71
Toaster	-	-	-	-	64.12 %	10
Cell phone	-	-	-	-	66.64 %	1
Total Images	197	662	344	102	734	-

Table 2. Detection confidence for each object class in the MIT Indoor Scenes dataset.

In Table 2, testing from the MIT Indoor Scenes dataset, determined the confidence scores of the detection model on indoor scenes images. Researchers selected 5 indoor scenes from the dataset and selected some object classes for the testing of confidence.

The result is separated into 5 columns for confidence 1 column for each indoor scene that describes confidence of a class per scene. The number of detections is also presented on the right most column. And the total number of images per indoor scene is added on the last row of the table.

The most common object detected on the dataset is the chair which has 77.14% confidence on the bathroom, 78.90% confidence on the bedroom, 73.88% confidence on the corridor, 80.56% confidence on garage and 81.36% confidence on the kitchen. While the least detected objects are umbrella, cellphone, and fork, and have a low confidence level. The least confidence level is the umbrella and the highest confidence level is the Table. Other objects with high count of detections are bottle, cup, table, vase, bowl, refrigerator and knife.

V. Conclusion and Recommendation

This study concludes the following:

1. The study created a tool that will detect and classify common objects with the combined use of pairwise Euclidean distance estimation that can give us information of the objects and estimated distance relation of objects in the image.
2. The study concludes that on dark low light images, the object detection has a high average confidence level on the objects detected, which means that the objects found on the dark images are likely correct.
3. The object detection model has a low accuracy on detecting small objects.
4. This study has a big potential in the Forensics field for crime scene investigation because of the image analysis and accuracy of the model which can provide a lot of useful information to the investigator.

For the recommendation:

Since this is an open source project, the source code can be found on the webpage, which can be useful to developers if they want to improve the project. For example an API can be created from the source code, and the project can be improved by adding features suited for crime scene detection such as:

- *Multiple file uploads*, this can be useful if there will be multiple images that need to be uploaded.
- *Video upload*, useful when finding certain objects on a CCTV feed and analyze different scenes from the video
- *Color search*, useful when searching for an item in the range of a specific color when looking for proofs.

The researchers recommend inputting more training data to improve the efficiency of the detection model in detection.

Based on what the researchers gathered, the confidence level of detecting images in low light areas. It is recommended to improve the model to train the detection on dark/low light images and on indoor areas.

As this project is meant for Crime Scene Investigation (CSI), it is also suggested to train object classes that are commonly found on crime scenes, specifically dangerous weapons. In addition, because of low confidence on the detection of small objects, It is recommended to train the model in detecting common small objects, which can be useful when finding some items like cellphones, cups and other small objects.

REFERENCES

[1] T. T. Contributor, "What is image? - Definition from WhatIs.com", *WhatIs.com*, 2021. [Online]. Available: https://whatis.techtarget.com/definition/image.

[2] R. Kapoor, "Importance of Still Photography at Scene of Crime: A Forensic vs. Judicial Perspective", *Crime-scene-investigator.net*, 2021. [Online]. Available: https://www.crime-scene-investigator.net/importance-of-still-photography-at-scene-of-crime.html.

[3] Crime Scene Investigation Laboratory Manual. Elsevier, 2014. doi: 10.1016/c2012-0-02802-5.

[4] "Equipment Needed for Crime Scene Investigation", *National Institute of Justice*, 2021. [Online]. Available: https://nij.ojp.gov/topics/articles/equipment-needed-crime-scene-investigation.

[5] "10 Modern Forensic Technologies Used Today - Forensics Colleges", *Forensics Colleges*, 2021. [Online]. Available: https://www.forensicscolleges.com/blog/resources/10-modern-forensic-science-technologies.

[6] E. R. Davies, "Image analysis in crime: progress, problems and prospects," presented at the IEE International Symposium on Imaging for Crime Detection and Prevention (ICDP 2005), 2005. doi: 10.1049/ic:20050079.

[7] S. Saikia, E. Fidalgo, E. Alegre, and L. Fernández-Robles, "Object Detection for Crime Scene Evidence Analysis Using Deep Learning," in Image Analysis and Processing - ICIAP 2017, Springer International Publishing, 2017, pp. 14–24. doi: 10.1007/978-3-319-68548-9_2.

[8] J. Osterburg and R. Ward, *Study Guide Criminal Investigation: A method for reconstructing the past*. 2004.

[9] C.-Y. Wang, H.-Y. Mark Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A New Backbone that can Enhance Learning Capability of CNN," presented at the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Jun. 2020. doi: 10.1109/cvprw50498.2020.00203.

[10] K. He, X. Zhang, S. Ren and J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 9, pp. 1904-1916, 1 Sept. 2015, doi: 10.1109/TPAMI.2015.2389824.

[11] S. Liu, D. Huang and Y. Wang, "Learning Spatial Fusion for Single-Shot Object Detection", *arXiv.org*, 2019. [Online]. Available: https://arxiv.org/abs/1911.09516.

[12] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement", *arXiv.org*, 2018. [Online]. Available: https://arxiv.org/abs/1804.02767.

[13] A. Bochkovskiy, C. Wang and H. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection", *arXiv.org*, 2020. [Online]. Available: https://arxiv.org/abs/2004.10934.

[14] "COCO - Common Objects in Context", *Cocodataset.org*. [Online]. Available: https://cocodataset.org.

[15] "MIT Indoor Scenes", *Kaggle.com*, 2019. [Online]. Available: https://www.kaggle.com/itsahmad/indoor-scenes-cvpr-2019.

[16] Y. Loh and C. Chan, "Getting to know low-light images with the Exclusively Dark dataset", *Computer Vision and Image Understanding*, vol. 178, pp. 30-42, 2019. Available: 10.1016/j.cviu.2018.10.010.