Proceedings of the International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT 2023)
IEEE Xplore Part Number: CFP23CV1-ART; ISBN: 978-1-6654-7451-1

2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT) | 978-1-6654-7451-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/IDCIoT56793.2023.10053440

Detection of Crime Scene Objects using Deep Learning Techniques

Nandhini T J
Department of Computer Science and Engineering
Saveetha School of Engineering
Saveetha Institute of Medical and Technical Science,
Chennai, India.
nandhinitj67@gmail.com

K Thinakaran
Department of Computer Science and Engineering
Saveetha School of Engineering
Saveetha Institute of Medical and Technical Science,
Chennai, India.
thinakarank.sse@saveetha.com

***Abstract*- Research on the detection of objects at crime scenes has flourished in the last two decades. Researchers have been concentrating on color pictures, where lighting is a crucial component, since this is one of the most pressing issues in computer vision, with applications spanning surveillance, security, medicine, and more. However, night time monitoring is crucial since most security problems cannot be seen by the naked eye. That's why it's crucial to record a dark scene and identify the things at a crime scene. Even when its dark out, infrared cameras are indispensable. Both military and civilian sectors will benefit from the use of such methods for night time navigation. On the other hand, IR photographs have issues with poor resolution, lighting effects, and other similar issues. Surveillance cameras with infrared (IR) imaging capabilities have been the focus of much study and development in recent years. This research work has attempted to offer a good model for object recognition by using IR images obtained from crime scenes using Deep Learning. The model is tested in many scenarios including a central processing unit (CPU), Google COLAB, and graphics processing unit (GPU), and its performance is also tabulated.**

Keywords: Object detection, Deep Learning, COLAB, CNN, Crime scene.

I. INTRODUCTION

In crime scene object detection, a wide range of computer vision applications, including autonomous driving, cutting-edge driving assistance systems, robotic visions, augmented reality, etc., make object identification a significant and active area of research. The primary goal of object detection is to locate and categorize particular items in still photographs and moving pictures.[1][2]. The process of focusing on an object involved in the vision process, such as visual tracking, human re-identification, and semantic segmentation, is typically seen as an essential step.[3] The semantic object detection technique makes use of several geometric patterns as proof to spot intriguing objects in pictures or movies.[4] The patterns of forms that display a comparable category of objects are used to train the object recognition models to distinguish between distinct categories. However, because the characteristics of basic item forms, object positions, and angles of view vary widely, it is challenging for a system to reliably identify every appearance of an object.[5] Multiple object detection uses similarities between the succession of photos or videos determine the movement of things. Target objects are first identified in multiple object detection and the technique is then followed to assess the itinerary of the items using the results of the detection. The journey of many objects is formed by semantic object detection with detection, which utilizes associated data from the existing track and fresh identification from each frame.[6] Thus, a sequence of detections with distinct identities is produced as a result of data association. Recognition of salient objects might be difficult when they have similar appearances. The moving objects are indicated for differentiating and monitoring the different objects in this situation.[7] When using a single moving camera, the movement of the global camera, which cannot be seen, contaminates the discernible indications of motion. Due to the decreased characteristics of images, such as blurriness in motion and defocus on films, which results in inconsistent categorization for comparable objects, object detection presented a challenge. Convolutional Neural Network (CNN), faster region-based CNN, spatial pyramid pooling network, region-based Fully CNN, You Only Look Once (YOLO), and Feature Pyramid Network (FPN) are some of the deep learning models that have been used to build the semantic object identification approach.[8]

This study develops an algorithm to categorise different crime scene photographs and to recognise different things in them. A video that is composed of photographs from crime scenes is taken into account and divided into frames. The frame rate range for video is between 45 to 120 frames per second, or 7200 pictures per minute[9][10]. When processing any video, this step is frequently taken. A seven-layered convolutional neural network is used to run the video after receiving the photos, and it typically recognises the images based on the trained images[11].

The current algorithms effectively identify items in some labeled photos, but they needed to be given positions, classes, and background distributions. However, when the objects were manually annotated, the assignment process was tedious and time-consuming. The handcrafted features of the old sliding window object identification approach had limitations that made it difficult to reliably recognize the items. Additionally, CNN succeeded in object detection by outperforming the conventional method. But because of difficult conditions including object occlusion, more fluctuation in object scale, and dim lighting, the CNN detector was not able to attain acceptable accuracy.

II. RELATED WORKS

This section reviews and describes existing object detection techniques, as well as their benefits and drawbacks. Sucheng Ren developed a new video crime scene object detection (VCSOD) method that uses a triple excitation mechanism. To solve the changing and contradicting Spatio-Temporal feature difficulties during the training phase, spatial and temporal excitations are offered. Additionally, semi-curriculum learning

is employed in training phase to initially reduce the task difficulty and get a higher level of convergence. In addition, we suggest the network's first online excitation during testing so that it can keep improving the saliency result by excitation while using the saliency map generated by the network. Extensive testing shows that our outcomes perform better than those of any competitors.[12]

Xuebin Qin developed an innovative deep network called U2-Net for the detection of salient objects. The main structure of our U2-Net is a two-level nested U-structure. No of the resolution, the network can gather greater local and global information from both shallow and deep layers because to the layered U-structure and our newly developed RSU blocks.[13]

Fu et.al. developed a region-based Convolutional Neural Network Framework for arbitrary and multi-scale item recognition in remote sensing pictures The feature fusion architecture was developed in order to extract detection characteristics based on the Region of Interest (RoI). To acquire the precise position of arbitrarily oriented objects, the Oriented Region Proposal Network (RPN-O) was constructed, and RoI pooling was utilised to avoid orientation changes. Because the established CNN architecture proved resistant to objects in remote sensing images, anchors with additional scales and angles were added to aid RPN in object detection[14]. However, the created feature fusion method was unable to identify comparable appearances and suffered from identifying the backdrop of images, which hampered object identification performance.

Cai and Vasconcelos created a cascade Region-based CNN approach for increasing the quality of object detection and segmentation. Inference uses the cascade to remove mismatched detectors and improve hypothesis quality. The resampling strategy increases hypothesis quality greatly by giving a positive training set with similar sizes for each detector and reducing overfitting. The created method, however, maximised the diversity of samples utilised to forecast the masked object because the segmentation procedure was a patch-based operation with a larger number of highly overlapping instances.[15]

Chen et al. propose leveraging reverse attention in the top-down pathway to steer residual saliency learning, which leads the network to uncover complementary object regions and details. However, the methods mentioned above use just individual resolution features in each decoder unit, which is insufficient for dealing with complicated and diverse scale challenges.[16]

Many recent papers in the literature use deep convolutional neural networks (e.g., AlexNet, GoogleNet, and VGG-Net) to detect and locate objects with class-specific bounding boxes. A CNN is typically made up of several convolution layers, followed by ReLU (Rectified Linear Units), pooling layers, and fully connected layers. The activations produced by a CNN's final layers can be used as a descriptor for object detection and classification.

The Faster R-CNN technique is used in a real-time system for crime scene evidence analysis that can find objects in an indoor environment. For object detection, the suggested system makes use of the Region Proposal Network and VGG-16 network. Compared to the current models, the suggested architecture provides a better level of accuracy.

III. PROPOSED CNN ARCHITECTURE

The suggested CNN architecture is displayed in Fig. 1 below. An infrared image measuring 640 by 480 pixels serves as the input. The architecture is made with the knowledge that the image is vulnerable to illumination, low resolution, and other influences, making it challenging for the image to effectively process and recognize the item. Therefore, this design consists of Seven Convolution layers with Max-pooling, One Flatten, and the SoftMax activation function. 32 filters are used in the first convolution layer, and 100 filters, each measuring 3 x 3, are used in the remaining six hidden layers. Max-pooling is also carried out with a scale of 2 x 2 at every layer. Following that, the convolutional layers are flattened and normalized using the SoftMax step. All layers employ the activation function "ReLu" and the output range is varying from 0 to infinity. The shear and stride values are both 1. The activation function of ReLu is given by the equation 1.

$$f(x) = \max(0, x) \quad (1)$$

Three different environments were used for the experimentation. The entire work was written in Keras and tested with two different datasets of 189 images and 147 images, which were later tested on GPU with 1820 images. The datasets under consideration are FLIR, and preliminary testing was performed on a system equipped with a Core i5, 8 GB RAM, and a 1 TB HDD. Later, experiments were carried out on a NIVIDIA DX-1 GPU with 128 GB RAM and a speed of 1 Tesla. The experiment was carried out with a dataset of 147 images and 189 images with 7 different classes and 21 different classes images and 27 images per class, respectively

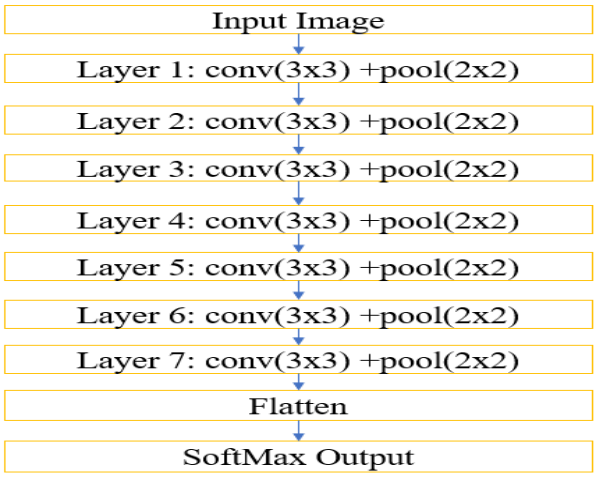

Fig. 1. Convolution Neural Network Architecture

The proposed model is evaluated based on various evaluation metrics such as Precision, Recall, F-Score, and Accuracy, which are defined as,

1) Precision = TP / (FP + TP)
2) Recall = TP / (FN + TP)
3) F1-Score = (2 * Recall * Precision) / (Recall + Precision)
4) Accuracy = (TP + TN) / (TP + FP + FN + TN)

TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives

A. TRAINING DATASET

Now that we have the data set, we need to pre-process it a little and label each of the images that were provided during training the data set. To do so, we can see that the names of each image in the training data set begin with "knife" or "gun," which we will exploit, and then we will use one hot encoder for the machine to understand the labels (Knife [1, 0] or gun [0, 1]). The datasets which have been used in this analysis is divided into two groups which is training and testing dataset in which 75% is taken for training and 25% for testing.

Fig.. 2. Dataset for Training

B. TESTING DATASET

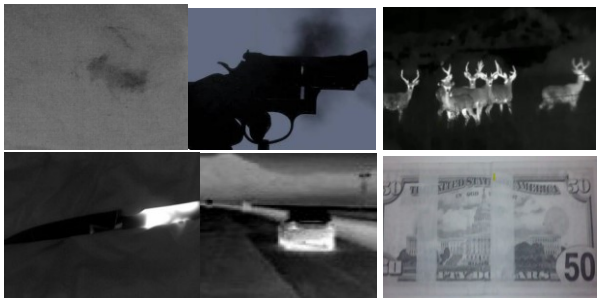

Fig.. 3. Dataset for Testing

IV. EXPERIMENTAL ANALYSIS

TABLE 1. CONFUSION MATRIX FOR 147 IMAGES EACH CLASS WITH 21 IMAGES

	PREDICTION						
objects	**Knife**	**Cell phone**	**Car**	**Animal**	**Gun**	**Blood**	**Currency**
Knife	27	0	0	0	0	0	0
Cell phone	0	26	0	0	0	0	0
Car	1	0	26	0	0	0	0
Animal	0	0	0	26	0	0	0
Gun	0	0	0	0	21	0	0
Blood	0	1	0	0	0	26	0
Currency	0	0	0	0	1	0	28

TABLE II. CONFUSION MATRIX FOR 189 IMAGES EACH CLASS WITH 27 IMAGES

	PREDICTION						
objects	**Knife**	**Cell phone**	**Car**	**Animal**	**Gun**	**Blood**	**Currency**
Knife	20	0	0	0	0	0	0
Cell phone	0	20	0	0	0	0	0
Car	1	0	21	0	0	0	0
Animal	0	0	0	21	0	0	0
Gun	0	1	0	0	20	0	0
Blood	0	0	0	0	0	20	0
Currency	0	1	0	0	1	0	21

TABLE III. ACCURACY FOR EACH CLASS FOR 147 IMAGES

Object	**Accuracy**
KNIFE	0.998
CELLPHONE	0.930
CAR	0.965
ANIMALS	0.964
GUN	0.972
BLOOD	0.975
CURRENCY	0.932

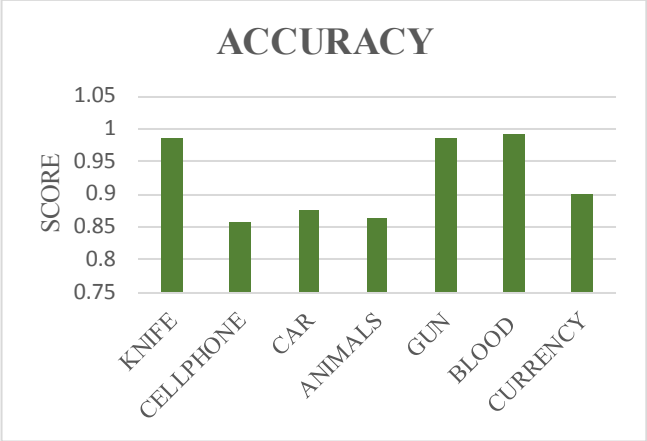

Fig.. 4. Accuracy for 147 Images in Each Class with 21 Images

The results are reported in the above table based on the accuracy discovered through testing on 147 photos divided into 7 classes, where the model accurately predicted with high accuracy for gun knife and blood. The analysis is carried out by using the google colab for predicting the accuracy of the data sets. The accuracy of the datasets is achived by increasing the training datasets than the testing datasets further the accuracy can also be improved if the amount of data in the training dataset is increased.

The following table provides the relevant confusion matrix for seven classes along with their TP (True Positives), TN (True Negatives), FP (False Positives), and FN (False Negatives).

TABLE IV. EVALUATION METRICS FOR 147 IMAGES IN EACH CLASS WITH 21 IMAGES

Experimental Epochs	Sensitivity	Specificity	Precision	F-Score
1	0.89	1	1	0.96
2	0.88	0.92	1	0.98
3	0.94	1	1	0.93
4	0.93	1	1	1
5	0.96	1	1	1
6	0.95	0.96	1	1
7	1	0.92	1	0.95

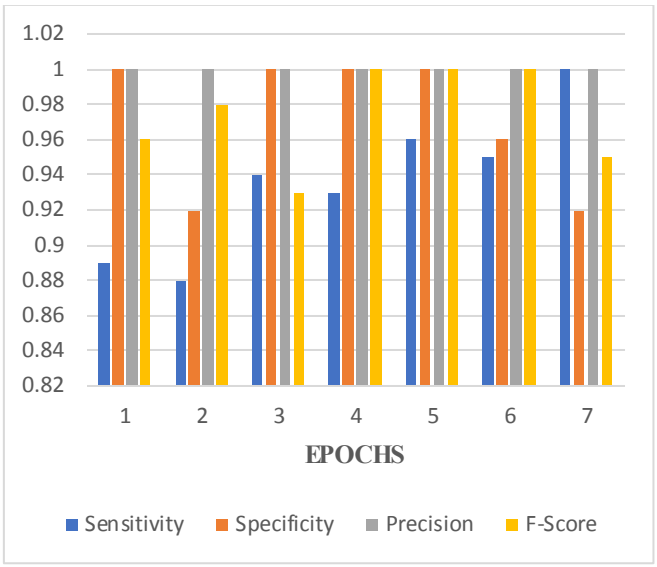

Fig.5. Evaluation Metrics for 147 Images In Each Class with 21 Images

TABLE V. ACCURACY FOR EACH CLASS FOR 189 IMAGES

Object	**Accuracy**
KNIFE	0.987
CELLPHONE	0.857
CAR	0.875
ANIMALS	0.863
GUN	0.985
BLOOD	0.993
CURRENCY	0.901

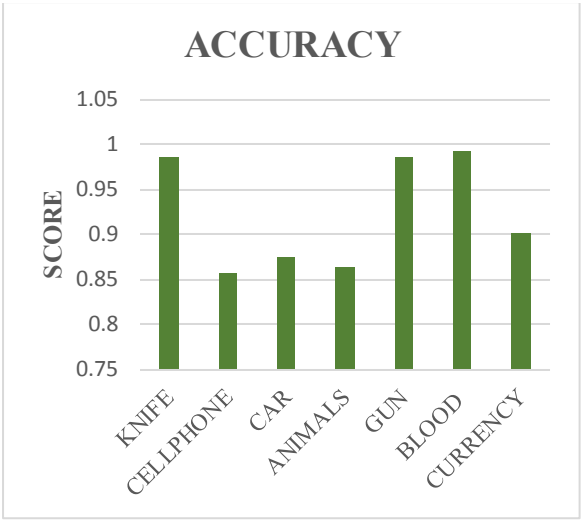

Fig. 6. Accuracy for 189 Images Each Class with 27 Images

TABLE VI. EVALUATION METRICS FOR 189 IMAGES IN EACH CLASS WITH 27 IMAGES

Experimental Epochs	Sensitivity	Specificity	Precision	F-Score
1	0.93	0.8	0.94	0.93
2	0.87	0.87	0.82	0.86
3	1	0.97	0.87	0.87
4	1	0.97	0.94	0.98
5	0.94	0.93	0.93	0.98
6	0.96	0.97	0.93	0.98
7	1	0.87	0.87	1

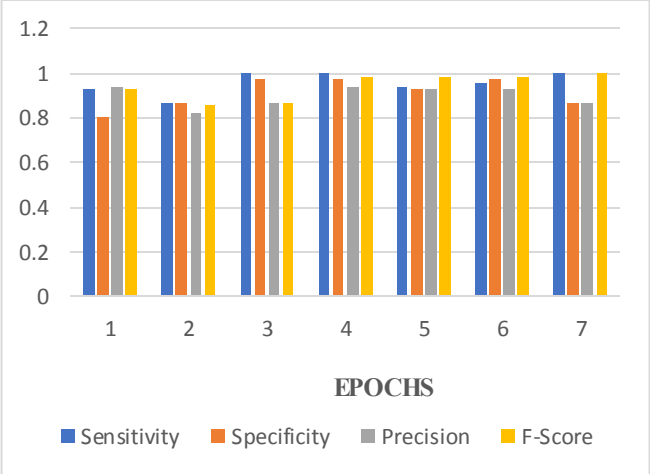

Fig.7. Evaluation Metrics for 189 Images in Each Class with 27 Images

TABLE VII. EXECUTION TIME FOR THREE ENVIRONMENTS

#Images	CPU(Mins)	COLAB (Mins)	GPU (Mins)
147	25	19	3.8
189	33	22	3.8

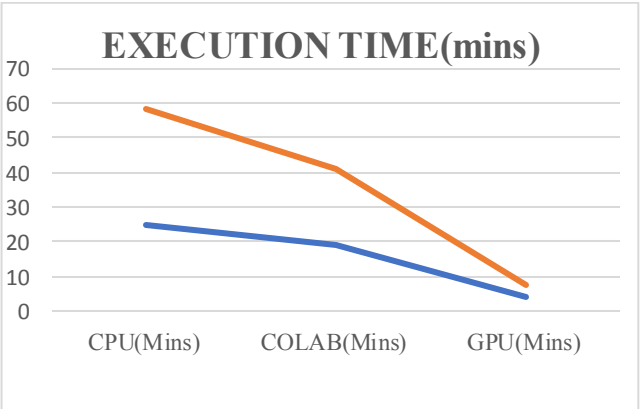

Fig. 8. Execution Time for three environments

V. CONCLUSION

The development of crime scene photographs will facilitate their use in numerous security and surveillance applications. According to the experimental results, the suggested CNN architecture outperforms in terms of accuracy, and using GPUs would cut processing time by 90%. As a result, the suggested model produced remarkably precise findings. However, by recognizing and identifying the object, there is still room for development. Additionally, it's crucial to comprehend the scene. The same effort can be expanded to find undersea objects that will be helpful for defence

REFERENCES

[1] A. Bathija, "Visual Object Detection and Tracking using YOLO and SORT," *Int. J. Eng. Res. Technol.*, vol. 8, no. 11, pp. 705–708, 2019, [Online]. Available: https://www.ijert.org

[2] M. Tiwari and R. Singhai, "A Review of Detection and Tracking of Object from Image and Video Sequences," *Int. J. Comput. Intell. Res.*, vol. 13, no. 5, pp. 745–765, 2017, [Online]. Available: http://www.ripublication.com

[3] T. Bergs, C. Holst, P. Gupta, and T. Augspurger, "Digital image processing with deep learning for automated cutting tool wear detection," *Procedia Manuf.*, vol. 48, pp. 947–958, 2020, doi: 10.1016/j.promfg.2020.05.134.

[4] J. Zhu, Z. Wang, S. Wang, and S. Chen, "Moving object detection based on background compensation and deep learning," *Symmetry (Basel).*, vol. 12, no. 12, pp. 1–17, 2020, doi: 10.3390/sym12121965.

[5] S. Ren, K. He, and R. Girshick, "Faster R-CNN : Towards Real-Time Object Detection with Region Proposal Networks," pp. 1–10.

[6] S. T. Naurin, A. Saha, K. Akter, and S. Ahmed, "A Proposed Architecture to Suspect and Trace Criminal Activity Using Surveillance Cameras," no. June, pp. 5–7, 2020.

[7] L. E. van Dyck, R. Kwitt, S. J. Denzler, and W. R. Gruber, "Comparing Object Recognition in Humans and Deep Convolutional Neural Networks—An Eye Tracking Study," *Front. Neurosci.*, vol. 15, no. October, pp. 1–15, 2021, doi: 10.3389/fnins.2021.750639.

[8] G. Kishore, G. Gnanasundar, and S. Harikrishnan, "A Decisive Object Detection using Deep Learning Techniques," *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, no. 1S, pp. 414–417, 2019, doi: 10.35940/ijitee.a1082.1191s19.

[9] S. A. T, S. R, V. R, and K. T, "Flying Object Detection and Classification using Deep Neural Networks," *Int. J. Adv. Eng. Res. Sci.*, vol. 6, no. 2, pp. 180–183, 2019, doi: 10.22161/ijaers.6.2.23.

[10] J. I.-Z. Chen and J.-T. Chang, "Applying a 6-axis Mechanical Arm Combine with Computer Vision to the Research of Object Recognition in Plane Inspection," *J. Artif. Intell. Capsul. Networks*, vol. 2, no. 2, pp. 77–99, 2020, doi: 10.36548/jaicn.2020.2.002.

[11] Y. He, X. Li, and H. Nie, "A Moving Object Detection and Predictive Control Algorithm Based on Deep Learning," *J. Phys. Conf. Ser.*, vol. 2002, no. 1, 2021, doi: 10.1088/1742-6596/2002/1/012070.

[12] S. Ren, C. Han, X. Yang, G. Han, and S. He, "TENet: Triple Excitation Network for Video Salient Object Detection," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12350 LNCS, pp. 212–228, 2020, doi: 10.1007/978-3-030-58558-7_13.

[13] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane, and M. Jagersand, "U2-Net: Going deeper with nested U-structure for salient object detection," *Pattern Recognit.*, vol. 106, 2020, doi: 10.1016/j.patcog.2020.107404.

[14] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.

[15] Z. Cai and N. Vasconcelos, "Cascade R-CNN: High quality object detection and instance segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 5, pp. 1483–1498, 2021, doi: 10.1109/TPAMI.2019.2956516.

[16] S. Chen, X. Tan, B. Wang, and X. Hu, "Reverse attention for salient object detection," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11213 LNCS, pp. 236–252, 2018, doi: 10.1007/978-3-030-01240-3_15.