iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** XI **Month of publication:** November 2024

DOI: https://doi.org/10.22214/ijraset.2024.65651

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XI Nov 2024- Available at www.ijraset.com

Evaluating the Impact of Various Existing Computer Vision Models on Reducing False Negatives and Positives in Commercial and Residential Security Surveillance Systems

Harish Balamurugan[1], Pranav Pakanati[2], Vedh Amara[3]
Heritage High School, Frisco, Texas, United States

Abstract: ***The following analytical study examines the contributions different computer vision models have made to reduce false negatives and positives in surveillance systems. Synthesizing the results of existing research regarding models in different environmental conditions, the work further compares the implications of these technologies at commercial and residential levels. The contexts of environmental and operational variances are tested apart to find any resultant trends and/or challenges. Conclusions are then derived from these results to recommend specific models that would perform better in a residential or commercial setting. Limitations are observed and results are discussed in a larger multi-disciplinary context.***
Keywords: Artificial intelligence, computer vision, machine learning, surveillance systems, object detection, false negatives, false positives

I. INTRODUCTION

Over the last decade, computer vision has been integrated into the architecture of security and surveillance systems, bringing numerous changes to the monitoring circumstances regarding public safety. Computer vision, with the development of Artificial Intelligence and machine learning, has enabled better analysis of video feeds, enhancing the efficiency and effectiveness of an entire surveillance operation. Before these models, traditional surveillance systems relied on a great deal of manual monitoring and basic motion detection techniques that often lead to high rates of false positives. Such inaccuracies compromise the reliability of a surveillance system and lead to higher security risks and inefficient operations. Computer vision models using convolutional neural networks, among other advanced algorithms, are purposed to recognize patterns or objects from either images or video streams with high accuracy. YOLO, Faster R-CNN, and DeepSort are prominent models; thus, it enables real-time object detection and tracking while notably reducing the false detection rate compared to earlier approaches. This evolution allows for even more responsive and intelligent surveillance systems that can be applied to commercial or residential settings. The effectiveness varies depending on the environmental conditions and operational contexts in which these models are put. This paper, therefore, seeks to assess different existing computer vision models for performance in reducing false negatives and false positives in commercial and residential security surveillance systems. This paper reviews the findings from existing research to compare the implications of such technologies within different contexts. This paper is essentially grounded on the hypothesis that, compared to residential settings, the use of computer vision models would more greatly decrease the number of both false negatives and positives in commercial environments. This hypothesis is derived from the understanding that commercial environments often have more structured settings, consistent lighting, and higher foot traffic-which may serve as a better background for effective model functionality.

II. LITERATURE REVIEW

There are various sets of algorithms and models that enable machines to interpret and understand visual information from the world. Among the most used models in object detection are YOLO, Faster R-CNN, and SSD.

YOLO: You Only Look Once is a real-time object detection algorithm which works very well when processing video streams. It divides an image into a grid and predicts bounding boxes and class probabilities directly from the full image in one evaluation, adding to its speed.

©IJRASET: All Rights are Reserved

This single-shot approach minimizes chances of missing objects—that is, false negatives—while typically achieving higher precision in detections. However, YOLO does have some weaknesses, such as struggles with smaller objects or those that are proximal to other similar objects.

Faster R-CNN: This new model outperforms its forerunners because it makes use of a region proposal network to predict bounding boxes, hence making the process of detection easier. Faster R-CNN boasts of accuracy in detection for various objects, which makes it suitable to be deployed in an environment that calls for precision of detection. It may, however, have a drawback in terms of speed compared to YOLO, which may be an essential factor in real-time surveillance systems.

SSD: Short for Single Shot Multibox Detector, SSD detects objects in a single pass; using a base convolutional network for feature extraction, and then using multiple convolutional layers to generate fixed-size bounding boxes at differing scales, which allows SSD to identify objects of varying sizes. This approach is also fast (in comparison to Faster R-CNN), simultaneously identifying and localizing objects.

Precision: It quantifies the quality of a model's positive predictions. High precision means that a large portion of the detected objects are actually true positives, something very important in security for avoiding redundant alerts.

Recall: This measures the model's performance concerning finding all relevant cases and calculates the ratio of true positives to the sum of actual positive cases. High recall means fewer false negatives, which is a desirable attribute to ensure that no valid threats are missed.

Average Precision (AP): AP is the standard metric in object detection tasks, computing precision-recall curves for each class and averaging scores. This gives a robust measure of the overall precision and recall of the model for various thresholds.

Mean Average Precision (mAP): mAP is a combination of the average precision and recall of each class averaged over all classes, which makes it a good predictor of the number of false positives and negatives that occur. Therefore, as mAP increases, the number of false positives and negatives are predicted to decrease.

Numerous studies have already investigated the performance of these computer vision models within surveillance systems. Research by (Pavithra S and B.Muruganantham) demonstrated that YOLO outperformed traditional detection methods in terms of speed and accuracy across multiple datasets, making it an attractive option for real-time applications. They noted a significant reduction in false positive rates, suggesting that its architecture could effectively minimize unnecessary alerts in surveillance contexts.

On the other hand, research by Kuo et al. (2018) investigates the feasibility of utilizing Mask R-CNN for retail product recognition as opposed to Faster R-CNN The results show that this network achieved an overall higher correct recognition rate, while recognizing the object in shorter time. Kristo et al. (2020) conducted a comparative performance study using different models, such as YOLO and R-CNN, on a dataset consisting of thermal images obtained from videos that simulated illegal terrorist and migrant movements in multiple weather conditions. They found that YOLOv3 performed significantly faster than other models along with higher accuracy. Liu et al. (2016) made a comparison between different versions of the Single-Shot Multibox Detector (SSD) and YOLO Models in detection of the growth of tomatoes in a greenhouse. Their purpose was to find the most appropriate model to be incorporated into the development of harvesting robots. The study found that the SSD Inception v2 specifically had the best performance compared to other variants of SSD and YOLO, successfully analyzing green and red tomatoes.

III. METHODOLOGY

The proposed study follows a comparative analysis framework for evaluating different computer vision models to reduce false negatives and false positives in surveillance systems. It emphasizes two significant settings—commercial environments, like shopping malls or retail stores, and residential settings, such as single-family homes or apartment complexes—where environmental contexts affect model effectiveness.

The following are some key steps that will be taken to conduct proper analysis:

1) Literature Review: A careful review of related work will identify the appropriate metrics and performance measures of each model, both in commercial and residential settings. The false positive and negative rates will be noted, together with metrics like precision, recall, and F1-scores to denote the mAP score for different conditions.
2) Comparative Analysis: Results from the literature review will be synthesized and compared across different studies. This will highlight performance discrepancies of models in the two settings and identify any contextual factors influencing these results. While we conduct an extensive analysis of computer vision models in surveillance, several limitations are also recognized:
3) Data Availability: Complex and quality surveillance model-specific datasets might not be readily available. Furthermore, some studies may not provide detailed statistics on performance, hence limiting the comparative analysis.

4) Heterogeneity of Study Design: Studies may utilize different methodologies, evaluation protocols, or definitions of success. Efforts will be made to standardize comparisons as far as possible.
5) Focus on Existing Research: The study primarily relies on existing literature and datasets, which limits the scope of original data collection. Future research could complement these findings with primary data collection and real-world experimentation.

Overall, to compare the false positive and negative rate of object detection models in residential and commercial contexts, how the models perform in the varying conditions between these contexts will be considered, using key metrics like mean average precision (mAP) which serves as a general indicator of both average precision and recall, which themselves are indicators of true positives and negatives respectively. An example of varying conditions that will be considered is the differences in lighting conditions, as residential contexts are more likely to have harsher lighting conditions compared to uniform lighting in commercial contexts. Another example is traffic, as models that have a higher frame rate and can track objects better over time while they are moving will miss less objects, decreasing their rates of false negatives.

IV. RESULTS

This section highlights the various performance metrics compiled from the literature review and existing studies about the use of YOLO, Faster R-CNN, and SSD to reduce false negatives and false positives in the environments encountered in residential and commercial settings.

A. Advantages and Disadvantages of Existing Methods for Object Detection

Faster R-CNN: While it is accurate and almost real-time taking about 0.12 seconds per image, it may not be fast enough to be used in true real-time applications, as security cameras would require.

YOLO: It is efficient at finding the location of objects, and can run very quickly in comparison to Faster R-CNN (making it viable for security cameras) especially using lower resolution input images; however, this comes at the cost of not being able to correctly identify smaller objects, which could prove to be a problem in residential contexts, where smaller objects are much more likely to be scattered throughout compared to the structured environment of a commercial setting. This could also prove to be a problem in retail, if smaller items are being sold on shelves.

SSD: With SSD requiring one network, it can tag and locate objects in one pass, making it much faster than Faster R-CNN and comparable to YOLO. However, this comes at a cost of being less accurate at detecting objects compared to Faster R-CNN and YOLO, albeit just slightly. Therefore, this model typically picks up more false negatives and positives compared to its counterparts.

B. Performance Metrics of Various Models in Varying Conditions

Table 1 summarizes the key performance outcomes for a variety of models, using the PASCAL VOC 2007 dataset (S A Sanchez et al., 2020)

Table 1

Detection Frameworks	mAP (0.0 - 100.0)	FPS
Fast R-CNN	70.0	0.5
Faster R-CNN VGG-16	73.2	7
Faster R-CNN ResNet	76.4	5
SSD300	74.3	46
SSD500	76.8	19
YOLO	63.4	45
YOLO v2 288 x 288	69.0	91
YOLO v2 352 x 352	73.7	81

Detection Frameworks	mAP (0.0 - 100.0)	FPS
YOLO v2 416 x 416	76.8	67
YOLO v2 480 x 480	77.8	59
YOLO v2 544 x 544	78.6	40

Table 2 summarizes the key performance outcomes for Faster R-CNN and SSD in challenging lighting conditions, on High Dynamic Range (HDR) and Standard Dynamic Range (SDR) data respectively (R. Mukherjee et al., 2021).

Table 2

Detection Frameworks	mAP (0.00- 1.00)
Faster R-CNN (HDR)	0.57
Faster R-CNN (SDR)	0.55
SSD 512 (HDR)	0.66
SSD 512 (SDR)	0.49
SSD 300 (HDR)	0.56
SSD 300 (SDR)	0.45

Table 3 summarizes the key performance outcomes for YOLO and Faster R-CNN on a dataset (BDK100K) that reflects clear and rainy weather conditions (Hnewa et al., 2020).

Table 3

Detection Frameworks	mAP (0.00 - 100.00)
Faster R-CNN (clear conditions)	44.45
Faster R-CNN (rainy conditions)	39.00
YOLO-V3 (clear conditions)	52.62
YOLO-V3 (rainy conditions)	49.39

Table 1 establishes baseline performance in normal conditions. With this data, faster R-CNN and SSD both have comparable mAP scores, with SSD having significantly higher framerate. While increasing the image input size for SSD increases mAP score, it significantly decreases the framerate, showing a potential trade-off that needs to be made in terms of efficiency and reducing false positives / negatives. YOLO v2 was the highest performing model considering both benchmarks, as higher input image sizes led to the highest mAP scores present in the study, while still maintaining a comparably high frame rate that often outperforms the frame rate of security cameras. Therefore, in nominal conditions, YOLO v2 is a comparably good model at reducing false negatives and positives and should be applicable to security cameras with comparatively more success to Faster R-CNN and SSD.

However, conditions will not always be nominal. For example, commercial environments are more likely to have consistent lighting, as is present in retail stores and larger warehouses. However, in a residential setting with more natural lighting, the overall lighting conditions can vary and be more challenging for an object detection model. An example of this would be light shafts coming through windows that can significantly affect the exposure of an image, therefore negatively impacting the false positive / negative rate of object detection. This is shown in Table 2, where both Faster R-CNN and SSD had relatively lower performance, as shown with the SDR dataset. However, on the HDR dataset, both models saw an increase in mAP scores, with SSD512 outperforming SSD300 and even had a higher score than Faster R-CNN. Therefore, in a situation where lighting conditions can be challenging, using HDR imagery can improve object detection capabilities regardless of what model is used, but HDR + SSD512 will likely provide the greatest benefits in a residential setting. It should be noted, however, that this study did not have results on YOLO, which leaves room for further research to be done and see its performance relative to SSD300, SSD512, and Faster R-CNN.

Furthermore, both residential and commercial environments can have suboptimal weather conditions. This makes performance in these conditions (i.e. rainy weather) important for both residential and commercial cameras and should be considered alongside performance in challenging lighting conditions specifically for enhanced residential performance. In Table 3, YOLO v3 has better mAP scores in comparison to Faster R-CNN, and this holds true in rainy conditions, where both models take a performance hit. Therefore, regardless of a residential or commercial setting, YOLO typically has better performance in terms of mAP scores (which is indicative of reduced false positives and negatives) and boasts significantly higher frame rates—ultimately making it a better fit for security cameras than Faster R-CNN.

However, when considering a residential setting, SSD may be a better fit than YOLO or Faster R-CNN. This is because as discussed earlier, it had higher performance on an HDR dataset compared to Faster R-CNN, which may come into play in a residential setting. Additionally, considering SSD's increased frame rate in comparison to Faster R-CNN, and YOLO's struggles with smaller objects, SSD may be a significantly better option. However, more research needs to be done regarding YOLO's performance in difficult lighting environments. If it is found that YOLO performs better than SSD in these environments, then it may be a comparable option to SSD in a residential environment.

Now, in a strictly commercial environment, although frame rate is less of a concern in smaller retail-like environments where most objects do not move around much (e.g. corner stores), Faster R-CNN's degraded performance in rainy conditions relative to YOLO makes it a worse option for security cameras, as there are going to be more false positives and negatives as shown with a lower mAP. In commercial environments that have higher traffic (commercial parking lots), Faster R-CNN becomes a worse option, as it may misidentify or fail to track objects that are moving faster, increasing the rate of false negatives. Overall, because of YOLO's comparable performance and slightly better performance in the rain, it should be a better option overall for commercial environments; although further research should be done on SSD's performance in the rain, and if it could be a better overall option than both YOLO and Faster R-CNN.

V. DISCUSSION

The findings from this study underscore how different computer vision models perform variably in commercial versus residential surveillance environments. While all models exhibit high mean average precision in settings common to both residential and commercial settings, the effectiveness of all models except for SSD declines significantly in residential settings, because of more challenging lighting conditions. In addition, YOLO does better in rainy weather conditions compared to R-CNN, and both YOLO and SSD have higher frame rates.

The results suggest that security practitioners and system designers must carefully evaluate the specific needs of the surveillance environment when selecting models. By acknowledging contextual features, they can optimize model performance to effectively address unique challenges.

For example, in a residential context, security cameras would benefit from using SSD compared to YOLO, because of YOLO's degraded performance with smaller objects and SSD's better performance in more challenging lighting conditions. However, SSD is only able to have better performance than Faster R-CNN in these challenging lighting conditions when it was trained on an HDR dataset; therefore, residential security camera designers should consider using a HDR or WDR (Wide Dynamic Range) camera for these gains to be fully realized. Another consideration to be made is frame rate. Considering YOLO and SSD have considerably higher frame rates than Faster R-CNN, if the security camera has a higher frame rate (i.e. 30fps, 60fps) then using R-CNN would be a bottleneck in the speed at which the camera could detect and track the locations of objects.

The theoretical contributions of this study advocate for further development in computer vision technologies to be contextually adaptive. Findings indicate that enhancing model flexibility can result in improved detection across varying environments. Developing algorithms capable of learning and adapting to specific environmental conditions could alleviate some performance issues observed in residential settings.

From a practical standpoint, these insights have significant implications for developers and users of surveillance systems. The study stresses the necessity of considering the operational environment during model selection and deployment. For commercial applications, investing in more sophisticated models like Faster R-CNN or YOLO may yield higher precision and recall rates, bringing down the rate of false negatives and positives. Conversely, in residential environments, employing a hybrid approach using both SSD and YOLO could allow for gains with harder lighting conditions while minimizing YOLO's inaccuracies with smaller objects.

While this research offers valuable insights, several limitations should also be addressed:

Data Variability: Reliance on existing literature and publicly available datasets can introduce variability in study design, model training, and evaluation protocols. Such disparities can affect the validity of comparisons, as not all studies employ identical methodologies or performance measures. Environmental Factors: Although an attempt has been made to categorize environmental factors influencing detection performance, real-life variability cannot be fully captured. Additional variables, such as seasonal changes, urban versus rural settings, and the presence of reflective surfaces, could further explain performance trends.

Focus on Specific Models: By concentrating solely on YOLO, Faster R-CNN, and SSD, the study may overlook advances in other emerging technologies or models in the computer vision realm. Exploring a broader range of approaches, including transformer-based models, could yield complementary insights into surveillance performance and flexibility.

With the findings and limitations identified in this study, steps could be taken to draw further conclusions:

Enhanced Model Training: Further studies could focus on methodologies to bolster model robustness, especially in residential settings, where conditions are more unpredictable. Techniques such as domain adaptation, transfer learning, and synthetic data generation might help models become more adept at navigating diverse environments.

Integration of Multi-Model Approaches: Research into hybrid models that integrate multiple computer vision algorithms could yield better performance across all scenarios. A super-performing surveillance system leveraging the strengths of various models could enhance adaptability to varied environmental contexts.

Real-World Testing: Empirical research deploying various models in daily life scenarios will solidify the findings of this study. Longitudinal analyses could provide meaningful insights into model performance over time and under varying conditions, informing better deployment strategies for security practitioners.

VI. CONCLUSION

This comprehensive review of the contributions of YOLO, Faster R-CNN, and SSD to reducing false negatives and positives in security surveillance systems has highlighted notable performance variations across commercial and residential contexts. The key insights of the analysis include:

Model Performance: YOLO typically had the best precision and frame rate, beating out Faster R-CNN especially in rainy situations, while SSD had superior performance with challenging lighting conditions, beating out Faster R-CNN.

Environmental Impact: Structured commercial environments tend to favor optimal model functionality, leading to fewer false negatives and positives. Conversely, the complexities present in residential areas adversely affect model performance due to fluctuating lighting, for one.

Need for Contextual Adaptability: The findings emphasize that surveillance model deployment must account for specific environmental features. Security practitioners should carefully consider contextual factors when selecting and implementing surveillance systems.

©IJRASET: All Rights are Reserved

There are several limitations inherent to this study that may impact its generalizability. The dependency on existing literature might introduce variability in performance metrics due to differences in study designs and evaluation protocols. Moreover, while efforts were made to consider contextual variables, the complexity of the real-world limits control over numerous influencing factors. Lastly, restricting the analysis to three computer vision models may underrepresented advancements in other emerging technologies within the fast-evolving field. These findings carry significant implications for both theoretical understanding and practical application. Security professionals must recognize the importance of model capabilities and environmental contexts that are critical for enhancing system efficacy. Developers should consider implementing hybrid system designs to address the varying challenges presented by different settings—favoring higher accuracy within commercial contexts while promoting adaptability in residential areas.

The type of objects being detected also plays a role in what model should be used. For a high stakes security environment (e.g. military), Faster R-CNN has proved to have high accuracy at detecting weapons (Verma et al., 2017).

Future inquiry should be motivated by the findings and limitations detailed in this study. Key areas of exploration may include:

Hybrid Model Development: Encouraging research into hybrid surveillance applications capable of combining multiple models to leverage distinct algorithmic strengths and improve performance in diverse environments.

Field Studies and Longitudinal Research: Conducting naturalistic studies examining computer vision model performance over time across various surveillance tasks can help refine optimal deployment strategies and inform best practices for system design.

Advanced Training Techniques: Investigating new training methodologies such as transfer learning and domain adaptation could significantly enhance model generalization capabilities across differing environments, particularly within residential settings where conditions are less predictable.

In conclusion, the potential of computer vision to enhance security via surveillance is vast, yet it must be approached thoughtfully. The study has illustrated the extent to which the environment influences model performance, indicating that careful selection and deployment strategies, informed by contextual understanding, are essential. Continued development and research will incrementally improve the efficacy of these systems, ultimately leading to increased safety and security in both commercial and residential applications.

By fostering collaboration and ongoing dialogue among researchers, developers, and practitioners, the evolving landscape of computer vision in surveillance can effectively address emerging challenges and maximize its protective capabilities.

REFERENCES

[1] Hnewa, Mazin, and Hayder Radha. "Object detection under rainy conditions for autonomous vehicles: A review of state-of-the-art and emerging R.

[2] Kristo, Mate, et al. “Thermal Object Detection in Difficult Weather Conditions Using YOLO.” IEEE Access, vol. 8, 2020, pp. 125459–125476, https://doi.org/10.1109/access.2020.3007481. Accessed 31 July 2020.

[3] Kuo, Matthew, et al. “Study on Mask R-CNN with Data Augmentation for Retail Product Detection.” 2021 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), 16 Nov. 2021, https://doi.org/10.1109/ispacs51563.2021.9651028. Accessed 23 Jan. 2023.

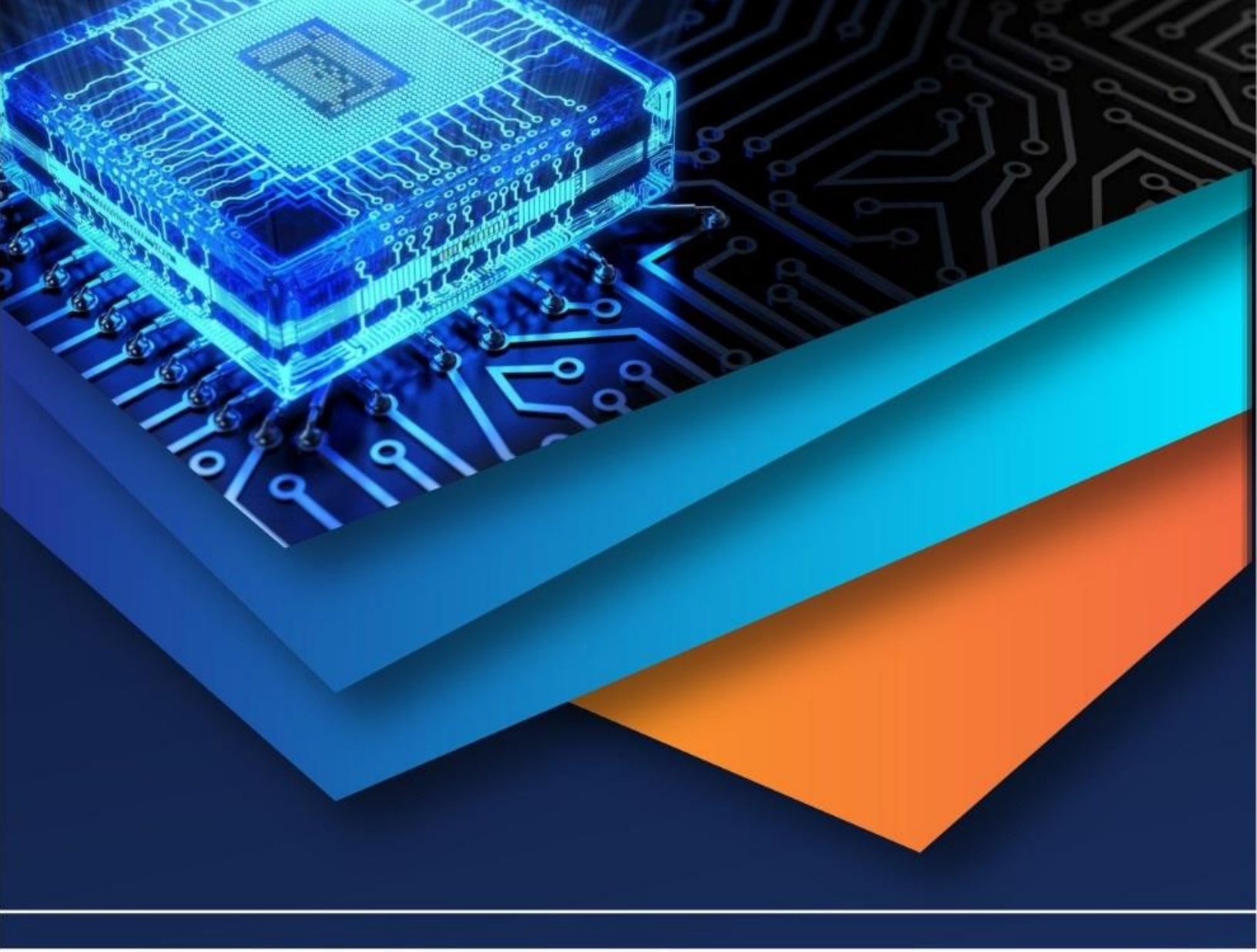
[4] Magalhães, Sandro Augusto, et al. “Evaluating the Single-Shot MultiBox Detector and YOLO Deep Learning Models for the Detection of Tomatoes in a Greenhouse.” Sensors, vol. 21, no. 10, 1 Jan. 2021, p. 3569, www.mdpi.com/1424-8220/21/10/3569, https://doi.org/10.3390/s21103569.

[5] Mukherjee, M. Bessa, P. Melo-Pinto, and A. Chalmers, "Object Detection Under Challenging Lighting Conditions Using High Dynamic Range Imagery," in IEEE Access, vol. 9, pp. 77771-77783, 2021, doi: 10.1109/ACCESS.2021.3082293.

[6] Pavithra S, and B. Muruganantham. “Panoramic Video Surveillance: An Analysis of Burglary Detection Based on YOLO Framework in Residential Areas.” Journal of Computer Science, vol. 19, no. 11, 1 Nov. 2023, pp. 1345–1358, https://doi.org/10.3844/jcssp.2023.1345.1358.

[7] Verma, Gyanendra K., and Anamika Dhillon. “A Handheld Gun Detection Using Faster R-CNN Deep Learning.” Proceedings of the 7th International Conference on Computer and Communication Technology - ICCCT-2017, 2017, https://doi.org/10.1145/3154979.3154988. Accessed 8 Nov. 2021.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)