Deep Learning-Based Real-Time WeaponDetection in CCTV Surveillance Footage

[1.] S. Tejaswi (Assistant Professor), [2.] Vikas, [3.] Tarani, [4.] Aswan, [5.] Apparao
Computer Science Engineering, Sanketika Vidya Parishad Engineering College, Visakhapatnam,India

Abstract- ***In today's modern world, ensuring security and safety is paramount for a country's economic strength and attracting investors and tourists. While Closed Circuit Television (CCTV) cameras are commonly used for surveillance, they still rely on human supervision to detect illegal activities such as robberies. The challenge remains in developing a system that can automatically detect such activities, especially weapon-related threats, in real time despite advancements in deep learning algorithms, hardware processing speed, and cameratechnology.***

This work focuses on enhancing security using CCTV footage to detect harmful weapons by leveraging state-of-the-art open-source deep learning algorithms. The approach involves binary classification, with pistols as the reference class, and introduces the concept of including relevant confusion objects to reduce false positives and false negatives. Due to the lack of a standard dataset for real-time scenarios, a custom dataset was created using weapon photos from various sources, including personal cameras, internet images, YouTube CCTV videos, GitHub repositories, data from the University of Granada, and the Internet Movies Firearms Database (IMFDB).

Two main approaches were employed: sliding window/classification and region proposal/object detection. Several deep learning algorithms, including VGG16, Inception-V3, Inception-ResnetV2, SSDMobileNetV1, Faster-RCNN Inception-ResnetV2 (FRIRv2), YOLOv3, and YOLOv4, were tested based on precision and recall, which are more crucial metrics than accuracy for object detection tasks.

Among these algorithms, YOLOv4 demonstrated superior performance, achieving an F1-score of 91% and a mean average precision of 91.73%, surpassing previous benchmarks. This highlights its effectiveness in accurately detecting harmful weapons in CCTV footage, contributing significantly to enhancing security measures.

I. INTRODUCTION

The global increase in crime rates, particularly involving handheld weapons during violent activities, poses a significant challenge to ensuring public safety and security. To foster economic growth, countries must maintain law and order, providing a peaceful environment for investment and tourism. However, the prevalence of firearms-related crimes remains critical in many regions worldwide, particularly in countries where firearm possession is legal. In today's interconnected world, the spread of information, whether true or false, through various media channels, including social media, can significantly influence individuals' behavior. This can lead to heightened levels of anxiety, decreased impulse control, and an increased susceptibility to radicalization, particularly when individuals have access to firearms.

Recent years have seen a rise in incidents involving harmful weapons in public areas, such as the attacks on mosques in New Zealand and active shooter incidents in the USA and Europe. These events highlight the urgent need for effective security measures to prevent such tragedies. Closed Circuit Television (CCTV) cameras play a crucial role in enhancing security by providing surveillance and evidence for crime investigations. Countries around the world have invested heavily in surveillance systems, with millions of cameras installed in public spaces.

However, traditional surveillance systems relying on human operators to monitor camera feeds have limitations, including reduced attention span and the inability to monitor multiple screens simultaneously. To address these challenges, there is a growing need for automated surveillance systems capable of detecting threats, such as weapons, in real-time and alerting security personnel promptly.

Despite advancements in technology, there has been limited research on algorithms for weapon detection in surveillance cameras, with most studies focusing on concealed weapon detection using X-rays or millimeter wave imaging. However, deep learning, particularly Convolutional Neural Networks (CNN),

has shown remarkable success in image-processing tasks, including object detection and localization.

This article presents a novel approach to automatic weapon detection and classification in real-time using state- of-the-art deep learning models. By focusing on pistols, revolvers, and other handheld weapons, the system aims to detect potential threats and alert operators and authorities promptly. The research involved creating a custom dataset comprising images extracted from CCTV videos, online repositories, and databases such as the Internet Movie Firearms Database (IMFDB). Deep learning models, trained using transfer learning and pre-trained models such as ImageNet and COCO, were evaluated for their performance in real-time weapon detection.

The main contributions of this work include:

A comprehensive study on weapon detection in real-time CCTV video streams, considering low-resolution andlow-brightness scenarios.

Identification of the most suitable CNN-based object detector for weapon detection.Development of a new dataset tailored for real-time detection.

Introduction of related confusion classes to reduce false positives and negatives.

Evaluation of various deep learning models, with YOLOv4 demonstrating the best performance in terms ofspeed and accuracy.

The remainder of the paper is structured as follows: Section II discusses related work, Section III explains the implementation methodology based on deep learning algorithms, Section IV details the dataset construction and preprocessing, Section V presents experimental results, and Section VI concludes the paper with discussions on future directions.

II.RELATED WORK

The problem of real-time detection and classification of objects, particularly in the context of surveillance using Closed Circuit Television (CCTV), has garnered increasing attention with advancements in CCTV technology, processing hardware, and deep learning models. While early efforts primarily focused on concealed weapon detection (CWD), recent research has shifted towards automated weapon detection using deep learning algorithms.

CWD techniques, initially developed for airport security and luggage control, relied on imaging methods such as millimeter-wave and infrared imaging. Various fusion-based techniques combining color visual and infrared images were proposed to enhance detection accuracy. However, these systems had limitations, including their reliance on metal detection, high costs, and health risks associated with X-ray scanners.

The evolution of CCTV technology has facilitated the application of automated image processing for public security purposes. Object detection algorithms have been widely used in surveillance systems for anomaly detection, deterrence, and human detection. However, the focus on firearm detection in CCTV footage has been relatively limited compared to other object detection tasks.

Early attempts at firearm detection in surveillance footage date back to 2007, with researchers proposing accurate pistol detection models using RGB images. However, these methods were not comprehensive in detecting various types of firearms in different scenes. Subsequent approaches utilized sliding window and region proposal algorithms, often incorporating Histogram of Oriented Gradient (HOG) models for feature extraction. While these methods achieved good accuracies, they were slow for real-time implementation.

In recent years, deep learning techniques, particularly Convolutional Neural Networks (CNNs), have shown promise in firearm detection. Researchers have explored various CNN architectures, including Faster R-CNN and YOLO (You Only Look Once), for real-time weapon detection in CCTV footage. Transfer learning approaches, leveraging pre-trained models such as ImageNet, have been employed to enhance detection accuracy.

Notable contributions include the development of novel datasets tailored for real-time detection, the introduction of related confusion classes to reduce false positives and negatives, and the evaluation of deep learning models for speed and accuracy in real-world scenarios. Recent studies have demonstrated significant improvements in detection accuracy, with some achieving precision and recall rates exceeding 90%.

Overall, while significant progress has been made in firearm detection using deep learning algorithms, challenges remain, particularly in real-time implementation and the development of robust datasets for training and testing. Future research efforts will likely focus on addressing these challenges

to further enhance the effectiveness of automated surveillance systems in ensuring public safety and security.

III.METHODOLOGY

Deep learning, inspired by the structure and functionality of the human brain, particularly artificial neural networks, serves as the foundation for the methodology employed in this work. Specifically, convolutional neural networks (CNNs) are leveraged due to their exceptional performance in object classification and detection tasks. Both classification and detection algorithms are utilized, with a focus on optimizing precision, speed, and localization accuracy for real-time implementation.

A. OBJECT RECOGNITION

Object recognition encompasses the processes of classification and localization, both of which are essential components of object detection. Classification involves predicting the category or class of an image, typically achieved using CNNs to extract features and assign probabilities to different classes. Localization determines the precise location of an object within an image, providing coordinates and dimensions for its bounding box. Object detection combines classification and localization to identify objects and their locations within an image.

B. CLASSIFICATION AND DETECTION APPROACH

Sliding Window/Classification Models: This approach involves sliding a window over the entire image to classify individual patches using an object recognition model. While exhaustive, this method is computationally intensive due to the need to search at multiple scales and aspect ratios.

Region Proposal/Object Detection Models: In contrast, region proposal methods generate bounding boxes for potential objects within an image, focusing computational resources on regions of interest. Techniques such as Selective Search and region-based CNNs (R-CNN) are used to generate region proposals efficiently.

C. TRAINING MECHANISM

Training deep learning models involves defining the problem, acquiring relevant datasets, preprocessing the data, and optimizing the model parameters using backpropagation and gradient descent algorithms. Training aims to minimize the loss function and improve the model's ability to generalize to unseen data.

D. CONFUSION OBJECT INCLUSION

To reduce false positives and negatives, relevant "confusion objects" are included in the dataset. These objects, such as wallets, cell phones, and metal detectors, resemble the target objects (e.g., pistols) and aid in improving overall accuracy and precision by distinguishing between similar classes.

E. CLASSIFIERS AND OBJECT DETECTORSSliding Window/Classification Models:

- VGG16
- InceptionV3
- Inception ResnetV2

Object Detectors for Real-Time Detection:

- SSD MobilNetV1
- YoloV3
- Faster RCNN-Inception ResNetV2
- YoloV4

IV. DATASET CONSTRUCTION, ANNOTATION AND PRE-PROCESSING (D-CAP)

The dataset construction process involves collecting relevant data, annotating images with labels and bounding boxes, and preprocessing the data for training. Dataset construction is critical for supervised learning, as the model's performance depends on the quality and diversity of the training data.

V.DATA PRE-PROCESSING AND ANNOTATION

In machine learning tasks, the quality and representation of the data significantly influence the performance of models. Data pre-processing plays a crucial role in enhancing the effectiveness of machine learning models by cleaning, standardizing, and extracting features from the dataset. The final training dataset is obtained after applying various pre-processing steps to the collected data.

Main Steps of Data Preprocessing:

Image Scaling: Ensuring uniform size or resolution of images in the dataset.Mean Normalization: Applying mean normalization to the images.

Image Labeling: Creating bounding boxes (annotation) on images to identify objects. This

involves storing the coordinates (x, y) and dimensions (width, height) of labeled objects in a suitable format such as XML, CSV, or TXT.

Image Filtering using OpenCV: Applying image filtering techniques using OpenCV library to enhance imagequality and reduce noise.

RGB to Grayscale: Converting images from RGB to grayscale format for simpler processing.
Equalization and Clahe: Applying histogram equalization and Contrast Limited Adaptive Histogram Equalization (CLAHE) techniques to improve image contrast and visibility.

VI.EXPERIMENTS, RESULTS AND ANALYSIS

The experiments involved real-time detection of weapons in CCTV streams under various conditions, including low resolution and low light. Previous work primarily focused on detecting high-quality images and videos, making it challenging to detect objects in real time under less favorable conditions. The performance of differentmodels trained and tested on the datasets mentioned in Table 1 was evaluated.

Results of Pre-processing Techniques:
Figures 1 and 2 depict the outcomes after applying the aforementioned pre-processing techniques to the images.

Figure 1

Evaluation Metrics:
The performance of the models was analyzed based on standard metrics such as:
F1-score: The harmonic-mean of precision and recall, providing a balanced measure of a model's accuracy.
Frame per Second (FPS): The rate at which frames are processed by the model, crucial for real-time applications.
Mean Average Precision (mAP): The average precision calculated across all classes, indicating the overall performance of the model.

These metrics were calculated using the following equations:
F1-score: F1 = 2 * (Precision * Recall) / (Precision + Recall)
Frame per Second (FPS): FPS = Total number of frames processed / Total time taken for processing
Mean Average Precision (mAP): Calculated based on the precision-recall curve for each class.
The analysis of results focused on comparing different approaches, such as the sliding window and region proposal methods, to address the real-time detection problem. Since pistols and revolvers accounted for a significant portion of weapons used in robbery cases, the evaluation considered datasets specifically tailored to this problem statement.

VII.CONCLUSION AND FUTURE WORK

In conclusion, this work presents a novel automatic weapon detection system for real-time monitoring and control. Implementing such a system is expected to significantly enhance security and improve law and order, particularly in regions plagued by violent activities. By providing a reliable means of detecting weapons in live CCTV streams, this system aims to mitigate security threats and create a safer environment for communities.
The implications of this work extend beyond security, as it has the potential to positively impact the economy byinstilling confidence among investors and attracting tourists who prioritize safety and security.

Key contributions of this work include:
Development of a new training database tailored for real-time scenarios. Training and evaluation of state-of-the-art deep learning models using both sliding window/classification and region proposal/object detection approaches.
Investigation of various algorithms to optimize precision and recall in weapon detection.
Through a series of experiments, it was observed that object detection algorithms with Region of Interest (ROI) outperformed those without ROI. Among the tested models, YOLOv4, trained on the new database, exhibited superior performance with minimal false positives and negatives. It achieved a mean average precision (mAP)of 91.73% and an F1-score of 91%, with a confidence score of nearly 99% across all types of images andvideos. These results indicate that

YOLOv4 effectively serves as an automatic real-time weapon detector, surpassing previous research efforts in terms of mAP and F1-score for real-time scenarios.

Future Work:

Despite the promising results, there remains room for improvement, particularly in reducing false positives and negatives. Future work will focus on further refining the precision and recall of the detection system. Additionally, there may be opportunities to expand the scope of the system by incorporating additional classes or objects of interest. However, the primary focus will be on enhancing the accuracy and reliability of weapon detection in real-time CCTV streams.

REFERENCES

[1] (2019). Christchurch Mosque Shootzings. Accessed: Jul. 10, 2019. [Online]. Available: https://en.wikipedia.org/wiki/Christchurch_mosque_shootings

[2] (2019). Global Study on Homicide. Accessed: Jul. 10, 2019. [Online]. Available: https://www.unodc.org/unodc/en/data-and-analysis/globalstudy-on-homicide.html

[3] W. Deisman, ‘‘CCTV: Literature review and bibliography,’’ in Research and Evaluation Branch, Community, Contract and Aboriginal Policing Services Directorate. Ottawa, ON, Canada: Royal Canadian Mounted, 2003.

[4] J. Ratcliffe, ‘‘Video surveillance of public places,’’ US Dept. Justice, Office Community Oriented Policing Services, Washington, DC, USA, Tech. Rep. 4, 2006.

[5] M. Grega, A. Matiolański, P. Guzik, and M. Leszczuk, ‘‘Automated detection of firearms and knives in a CCTV image,’’ Sensors, vol. 16, no. 1, p. 47, Jan. 2016.

[6] TechCrunch. (2019). China’s CCTV Surveillance Network Took Just 7 Minutes to Capture BBC Reporter.
Accessed: Jul. 15, 2019. [Online]. Available: https://techcrunch.com/2017/12/13/china-cctv-bbc-reporter/

[7] N. Cohen, J. Gattuso, and K. MacLennan-Brown. CCTV Operational Requirements Manual 2009. St Albans, U.K.: Home Office Scientific Development Branch, 2009.

[8] G. Flitton, T. P. Breckon, and N. Megherbi, ‘‘A comparison of 3D interest point descriptors with application to airport baggage object detection in complex CT imagery,’’ Pattern Recognit., vol. 46, no. 9, pp. 2420–2436, Sep. 2013.

[9] R. Gesick, C. Saritac, and C.-C. Hung, ‘‘Automatic image analysis process for the detection of concealed weapons,’’ in Proc. 5th Annu. Workshop Cyber Secur. Inf. Intell. Res. Cyber Secur. Inf. Intell. Challenges Strategies (CSIIRW), 2009, p. 20.

[10] R. K. Tiwari and G. K. Verma, ‘‘A computer vision-based framework for visual gun detection using Harris interest point detector,’’ Procedia Comput. Sci., vol. 54, pp. 703–712, Aug. 2015.

[11] R. K. Tiwari and G. K. Verma, ‘‘A computer vision-based framework for visual gun detection using SURF,’’ in Proc. Int. Conf. Electr., Electron., Signals, Commun. Optim. (EESCO), Jan. 2015, pp. 1–5.

[12] Z. Xiao, X. Lu, J. Yan, L. Wu, and L. Ren, ‘‘Automatic detection of concealed pistols using passive millimeter wave imaging,’’ in Proc. IEEE Int. Conf. Imag. Syst. Techn. (IST), Sep. 2015, pp. 1–4.

[13] D. M. Sheen, D. L. Mcmakin, and T. E. Hall, ‘‘Three-dimensional millimeter-wave imaging for concealed weapon detection,’’ IEEE Trans. Microw. Theory Techn., vol. 49, no. 9, pp. 1581–1592, Sep. 2001.

[14] Z. Xue, R. S. Blum, and Y. Li, ‘‘Fusion of visual and IR images for concealed weapon detection,’’ in Proc.5th Int. Conf. Inf. Fusion, vol. 2, Jul. 2002, pp. 1198–1205.

[15] R. Blum, Z. Xue, Z. Liu, and D. S. Forsyth, ‘‘Multisensor concealed weapon detection by using a multiresolution mosaic approach,’’ in Proc. IEEE 60th Veh. Technol. Conf. (VTC-Fall), vol. 7, Sep. 2004, pp. 4597–4601.

[16] E. M. Upadhyay and N. K. Rana, ‘‘Exposure fusion for concealed weapon detection,’’ in Proc. 2nd Int. Conf. Devices, Circuits Syst. (ICDCS), Mar. 2014, pp. 1–6.

[17] R. Maher, ‘‘Modeling and signal processing of acoustic gunshot recordings,’’ in Proc. IEEE 12th Digit. Signal Process. Workshop 4th IEEE Signal Process. Educ. Workshop, Sep. 2006, pp. 257–261.

[18] A. Chacon-Rodriguez, P. Julian, L. Castro, P. Alvarado, and N. Hernandez, ‘‘Evaluation of gunshotdetection algorithms,’’ IEEE

Trans. Circuits Syst. I, Reg. Papers, vol. 58, no. 2, pp. 363–373, Feb. 2011.

[19] (2019). From Edison to Internet: A History of Video Surveillance. Accessed: Jun. 13, 2019. [Online]. Available: https://www. business2community.com/tech-gadgets/from-edison-to-internet-ahistory-of-video- surveillance-0578308

[20] (2019). Infographic: History of Video Surveillance—IFSEC Global | Security and Fire News and Resources. Accessed: Sep. 15, 2019. [Online]. Available: https://www.ifsecglobal.com/video-surveillance/infographichistory-of-video-surveillance/

[21] W. Hu, T. Tan, L. Wang, and S. Maybank, ‘‘A survey on visual surveillance of object motion and behaviors,’’ IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 34, no. 3, pp. 334–352, Aug. 2004.

[22] A. C. Sankaranarayanan, A. Veeraraghavan, and R. Chellappa, ‘‘Object detection, tracking and recognition for multiple smart cameras,’’ Proc. IEEE, vol. 96, no. 10, pp. 1606–1624, Oct. 2008.

[23] S. Zhang, C. Wang, S.-C. Chan, X. Wei, and C.-H. Ho, ‘‘New object detection, tracking, and recognition approaches for video surveillance over camera network,’’ IEEE Sensors J., vol. 15, no. 5, pp. 2679–2691, May 2015.

[24] J. C. Nascimento and J. S. Marques, ‘‘Performance evaluation of object detection algorithms for video surveillance,’’ IEEE Trans. Multimedia, vol. 8, no. 4, pp. 761–774, Aug. 2006.

[25] N. Dalal and B. Triggs, ‘‘Histograms of oriented gradients for human detection,’’ Tech. Rep., 2005.

[26] C. Anagnostopoulos, I. Anagnostopoulos, G. Tsekouras, G. Kouzas, V. Loumos, and E. Kayafas, ‘‘Using sliding concentric windows for license plate segmentation and processing,’’ in Proc. IEEE Workshop Signal Process. Syst. Design Implement., Nov. 2005, pp. 337–342.

[27] M. Grega, S. Lach, and R. Sieradzki, ‘‘Automated recognition of firearms in surveillance video,’’ in Proc. IEEE Int. Multi-Disciplinary Conf. Cognit. Methods Situation Awareness Decis. Support (CogSIMA), Feb. 2013, pp. 45–50.

[28] I. Darker, A. Gale, L. Ward, and A. Blechko, ‘‘Can CCTV reliably detect gun crime?’’ in Proc. 41st Annu. IEEE Int. Carnahan Conf. Secur. Technol., Oct. 2007, pp. 264–271.

[29] I. T. Darker, A. G. Gale, and A. Blechko, ‘‘CCTV as an automated sensor for firearms detection: Human- derived performance as a precursor to automatic recognition,’’ Proc. SPIE, vol. 7112, Oct. 2008, Art. no. 71120V.

[30] I. T. Darker, P. Kuo, M. Y. Yang, A. Blechko, C. Grecos, D. Makris, J.-C. Nebel, and A. Gale, ‘‘Automation of the CCTV-mediated detection of individuals illegally carrying firearms: Combining psychological and technological approaches,’’ Proc. SPIE, vol. 7341, Apr. 2009, Art. no. 73410P.

[31] R. Al-Rfou et al., ‘‘Theano: A Python framework for fast computation of mathematical expressions,’’ 2016, arXiv:1605.02688. [Online]. Available: http://arxiv.org/abs/1605.02688

[32] F. Chollet. (2019). Fchollet. Accessed: Apr. 10, 2019. [Online]. Available: https://github.com/fchollet

[33] K. He, X. Zhang, S. Ren, and J. Sun, ‘‘Deep residual learning for image recognition,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.

[34] (2016). Weapon Detection in Surveillance Cameras Images. Accessed: Feb. 13, 2021. [Online]. Available: http://www.divaportal.org/smash/record.jsf?pid=diva2%3A1054902&dswid=-1974

[35] M. Nakib, R. T. Khan, M. S. Hasan, and J. Uddin, ‘‘Crime scene prediction by detecting threatening objects using convolutional neural network,’’ in Proc. Int. Conf. Comput., Commun., Chem., Mater. Electron. Eng. (IC4ME2), Feb. 2018, pp. 1–4.

[36] G. K. Verma and A. Dhillon, ‘‘A handheld gun detection using faster RCNN deep learning,’’ in Proc. 7th Int. Conf. Comput. Commun. Technol. (ICCCT), 2017, pp. 84–88.

[37] R. Olmos, S. Tabik, and F. Herrera, ‘‘Automatic handgun detection alarm in videos using deep learning,’’ Neurocomputing, vol. 275, pp. 66–72, Jan. 2018.

[38] J. Iqbal, M. A. Munir, A. Mahmood, A. Rafaqat Ali, and M. Ali, ‘‘Orientation aware object detection with application to firearms,’’ 2019, arXiv:1904.10032. [Online]. Available: https://arxiv.org/abs/1904.10032

[39] J. L. S. González, C. Zaccaro, J. A. Álvarez-García, L. M. S. Morillo, and F. S. Caparrini, ‘‘Real-time gun detection in CCTV: An open

problem,'' Neural Netw., vol. 132, pp. 297–308, Dec. 2020.
[40] (2017). Convolutional Neural Networks. Accessed: Aug. 15, 2018. [Online]. Available: http://cs231n.github.io/convolutional-networks/
[41] K. Simonyan and A. Zisserman, ''Very deep convolutional networks for large-scale image recognition,'' 2014, arXiv:1409.1556. [Online]. Available: http://arxiv.org/abs/1409.1556
[42] (2019). VGG16—Convolutional Network for Classification and Detection. Accessed: Dec. 19, 2018. [Online]. Available: https://neurohive. io/en/popular-networks/vgg16/
[43] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, ''Rethinking the inception architecture for computer vision,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 2818–2826.
[44] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, ''Inception-v4, inception-resnet and the impact of residual connections on learning,'' in Proc. 31st AAAI Conf. Artif. Intell., 2017, pp. 1–7.
[45] Medium. (2019). A Simple Guide to the Versions of the Inception Network. Accessed: Jul. 27, 2019. [Online]. Available: https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network- 7fc52b863202
[46] D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, ''SSD: Single shot multibox detector,'' in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, 2016, pp. 21–37.
[47] Medium. (2019). Understanding SSD MultiBox—Real-Time Object Detection in Deep Learning. Accessed: Aug. 19, 2019. [Online]. Available: https://towardsdatascience.com/understanding-ssd-multiboxreal- time-object-detection-in-deep-learning-495ef744fab
[48] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, ''MobileNets: Efficient convolutional neural networks for mobile vision applications,'' 2017, arXiv:1704.04861. [Online]. Available: http://arxiv.org/abs/1704.04861
[49] J. Redmon and A. Farhadi, ''YOLOv3: An incremental improvement,'' 2018, arXiv:1804.02767. [Online].
Available: http://arxiv.org/abs/ 1804.02767
[50] S. Ren, K. He, R. Girshick, and J. Sun, ''Faster R-CNN: Towards real-time object detection with region proposal networks,'' in Proc. Adv. Neural Inf. Process. Syst., 2015. pp. 91–99.
[51] (2019). Faster R-CNN Explained. Accessed: Aug. 25, 2019. [Online]. Available: https://medium. com/@smallfishbigsea/faster-r-cnn-explained-864d4fb7e3f
[52] Medium. (2019). Faster RCNN Object detection. Accessed: Aug. 27, 2019. [Online]. Available: https://towardsdatascience.com/faster-rcnn-object-detection-f865e5ed7fc4
[53] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, ''YOLOv4: Optimal speed and accuracy of object detection,'' 2020, arXiv:2004.10934. [Online]. Available: http://arxiv.org/abs/2004.10934
[54] GeeksforGeeks. (2020). Object Detection Vs Object Recognition Vs Image Segmentation. Accessed: Dec. 28, 2020. [Online]. Available: https://www.geeksforgeeks.org/object-detection-vs-object-recognitionvs-image- segmentation/
[55] (2019). ImageNet. Accessed: Jun. 5, 2019. [Online]. Available: http:// www.image-net.org/
[56] J. O. Laguna, A. G. Olaya, and D. Borrajo, ''A dynamic sliding window approach for activity recognition,'' in Proc. Int. Conf. User Modeling, Adaptation, Personalization. Berlin, Germany: Springer, 2011, pp. 219–230.
[57] J. Hosang, R. Benenson, P. Dollar, and B. Schiele, ''What makes for effective detection proposals?'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 4, pp. 814–830, Apr. 2016.
[58] R. Girshick, J. Donahue, T. Darrell, and J. Malik, ''Rich feature hierarchies for accurate object detection and semantic segmentation,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 580–587.
[59] A. Consulting. (2019). Selective Search for Object Detection (C++ / Python) | Learn OpenCV. Accessed: May 25, 2019. [Online]. Available: https://www.learnopencv.com/selective-search-for-object-detection-cpp- python/
[60] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, ''Gradient-based learning applied to document recognition,'' Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.