LIVE CCTV OBJECT DETECTION/TRACKING AND STORAGE OF EXPRESSIVE FOOTAGE: A SURVEY

Ankit Sharma[1], Arnav Zutshi[2], Mridul Sharma[3] ,Anshika Sharma[3], Prof. Pramila M. Chawan[3]

[1, 2, 3, 4]BTech Student, Dept. of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India
[5]Associate Professor, Dept. of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India

---

Abstract - *This project focuses on enhancing CCTV storage efficiency in traffic cameras by selectively recording relevant data. Leveraging YOLOv3 pre-trained weights, the system calculates frame-to-frame distance to identify significant changes, preventing storage loss. Integration with OpenCV's DNN module facilitates quick loading of YOLOv3 weights, enabling real-time object detection and classification. The proposed model not only accurately identifies objects causing changes but also encloses them with precise bounding boxes, offering a streamlined solution for efficient traffic surveillance.*

***Key Words*: YOLOv3, object detection, OpenCV**

1.INTRODUCTION

The surge in data from CCTV cameras, particularly in traffic surveillance, necessitates innovative solutions for efficient storage and real-time analysis. This survey paper explores the integration of YOLOv3, an advanced object detection algorithm, to address these challenges. The primary objectives include selectively storing relevant data to prevent storage loss and leveraging YOLOv3 for accurate object detection and classification.

By utilizing YOLOv3 pre-trained weights and a distance-based frame analysis, the system identifies significant changes between frames, ensuring the targeted recording of pertinent data. Integration with OpenCV's DNN module expedites YOLOv3 weight loading and enables real-time object classification. The proposed model enhances the analysis by enclosing detected objects with precise bounding boxes, contributing to a nuanced understanding of the surveillance scene.

This survey aims to provide a succinct yet comprehensive overview of recent advancements in optimizing data management and refining object classification in traffic camera surveillance systems.

1.1 YOLOv3

In the landscape of computer vision, YOLOv3 (You Only Look Once version 3) has emerged as a prominent force, particularly in real-time object detection. This survey paper endeavors to offer a comprehensive yet accessible exploration of YOLOv3, encompassing its architectural intricacies, training paradigms, and diverse applications across domains.

YOLOv3's distinctive grid-based methodology, simultaneously predicting bounding boxes and class probabilities for each grid cell, distinguishes it from conventional object detection approaches. Through an in-depth analysis, this survey illuminates the strengths, limitations, and the impact of YOLOv3's architectural choices on overall performance.

Beyond the technical nuances, the paper sheds light on the practical applications of YOLOv3 in domains such as surveillance, autonomous vehicles, and robotics. It also addresses challenges intrinsic to YOLOv3, including considerations for small object detection and the delicate balance between computational speed and detection accuracy.

In summary, this survey aspires to be a valuable resource, providing researchers, practitioners, and enthusiasts with a nuanced understanding of YOLOv3 and its evolving role in shaping the landscape of object detection within computer vision.

1.2 Cosine Similarity

Cosine similarity is a fundamental metric in textual analysis, known for its simplicity and effectiveness in measuring text vector similarity. This concise survey provides a brief exploration of cosine similarity, touching upon its theoretical foundations and practical applications in various domains. Aimed at researchers and practitioners, this overview offers quick insights into the role of cosine similarity in textual analysis..

2. Literature Review

2.1 COCO Dataset for Training:

The Common Objects in Context (COCO) dataset, introduced by Lin et al. in 2014, stands as a pivotal and widely adopted resource for training object detection models. COCO's significance lies in its comprehensive nature, encapsulating a diverse array of object categories within complex scenes. Its meticulous curation includes images with multi-object instances, diverse backgrounds, and intricate contextual variations, making it a rich repository for training algorithms to navigate real-world complexities. The dataset's expansive nature, with over 200,000 labeled images and more than 80 object categories, ensures that models trained on COCO exhibit a remarkable adaptability to a wide spectrum of scenarios. This diversity is especially crucial for enhancing the robustness and generalization capabilities of object detection algorithms, such as YOLOv3. The widespread adoption of COCO as a benchmark dataset in the computer vision community underscores its pivotal role in advancing the field, serving as a cornerstone for fostering innovation and pushing the boundaries of object detection research and development.

2.2 OpenCV in YOLOv3 Integration:

The Open Source Computer Vision Library (OpenCV), introduced by Bradski in 2000, stands as a cornerstone in the seamless integration of YOLOv3 into object detection frameworks. Specifically, OpenCV's DNN (Deep Neural Networks) module assumes a pivotal role in this synergy. By facilitating the swift loading of YOLOv3 pre-trained weights, the DNN module significantly optimizes the efficiency of real-time object detection in practical surveillance scenarios. This integration not only exemplifies the versatility of OpenCV but also underscores its instrumental contribution to the practical deployment and performance enhancement of YOLOv3 in dynamic surveillance environments.

2.3 Neural Networks for Object Classification:

Neural Networks, pioneered by LeCun et al. (2015), stand as a transformative force in refining object classification accuracy. This groundbreaking development leverages intricate architectures to emulate human brain functions for learning and decision-making.

Within YOLOv3, Convolutional Neural Networks (CNNs) play a pivotal role. Mimicking the human visual processing system, CNNs excel in extracting hierarchical features from images, enabling precise pattern recognition. Integrated into YOLOv3, CNNs enhance the model's ability to discern intricate visual information, contributing significantly to reliable object classification. This synergy allows YOLOv3 to excel in real-time object detection, particularly in dynamic environments, demonstrating the prowess of Neural Networks in advancing object classification capabilities.

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

2.4 Cosine Similarity:

cosine_similarity(A, B) = (A . B) / ||A|| ||B||
Where:
A and B are vectors is the dot product operator || || is the norm operator
YOLOv3 Bounding Box Coordinates:
P_c, b_x, b_y, b_w, b_h

Where:
P_c is the objectness score
b_x and b_y are the coordinates of the bounding box center
b_w and b_h are the width and height of the bounding box, respectively

These formulas are used to calculate the cosine similarity between two vectors and to identify the coordinates of a bounding box around an object. The cosine similarity is used to measure the similarity between two images, while the bounding box coordinates are used to identify the location and size of an object in an image.

The convergence of these frameworks represents a holistic approach to address the challenges in traffic camera surveillance. YOLOv3's real-time capabilities, enriched by the diverse training data from COCO, showcase the efficacy of this combination. OpenCV's DNN module further streamlines the operationalization of YOLOv3, enhancing its applicability in real-world scenarios. The incorporation of Neural Networks elevates object classification accuracy, providing a comprehensive solution for the nuanced demands of optimizing data storage and object detection in traffic surveillance systems.

3. PROPOSED SYSTEM

3.1 Problem Statement

In the realm of CCTV surveillance, particularly traffic cameras, the challenge lies in efficiently managing storage resources and detecting relevant changes. The project addresses this predicament by aiming to selectively record pertinent data while identifying and classifying the objects causing significant alterations in the video frames.

3.2 Problem Elaboration

In the context of Live-CCTV, the proposed problem statement addresses the need for an advanced system to optimize data storage in traffic camera surveillance, building upon the existing project's foundation.

<u>Current Scenario:</u>
Currently, the project employs YOLOv3 and OpenCV to detect relevant changes in a live CCTV feed, preventing unnecessary data storage. It calculates the distance

between frames and classifies objects causing changes, showcasing effectiveness in real-time surveillance.

Proposed Enhancement:
The proposed problem statement aims to elevate the project's capabilities further. The focus lies on refining motion tracking and classification mechanisms using cutting-edge techniques. The key enhancements include:

Feature Vector Integration:
Incorporating Convolutional Neural Networks (CNNs) to extract detailed feature vectors.

Employing these feature vectors alongside Euclidean distance calculations for more accurate motion tracking.

Real-Time Car Tracking:
Evolving the system into a real-time car tracker.

Implementing the identification of vehicles and saving their number plates for subsequent tracking in multiple CCTV feeds.

Self-Driving Cars Integration:
Exploring the potential application of the project in self-driving car technologies.

Investigating how the system can contribute to real-world scenarios in autonomous vehicles.

3.3 Proposed Methodology

The proposed methodology for the Live-CCTV project revolves around refining data storage efficiency and enhancing real-time object detection in traffic camera surveillance. The approach combines computer vision concepts, Deep Learning, and the integration of advanced algorithms. Here is a step-by-step breakdown of the methodology:

Frame-to-Frame Distance Calculation:
Utilize Euclidean distance to calculate the dissimilarity between two consecutive frames of the video feed.

Establish a threshold to determine the significance of the detected change. Frames exceeding this threshold are considered relevant.

Object Detection using YOLOv3:
Employ YOLOv3's pre-trained weights and configuration file for efficient and accurate object detection.

Leverage the grid-based methodology of YOLOv3 to simultaneously predict bounding boxes and class probabilities for each grid cell in the frame.

Motion Tracking with Cosine Similarity:
Use Cosine Similarity to measure the similarity between two consecutive frames.

Apply a threshold to the cosine similarity score to identify and track significant motion or changes between frames.

Integration with OpenCV's DNN Module:
Utilize OpenCV's DNN module to seamlessly load YOLOv3 pre-trained weights.

Leverage the optimized functionality of OpenCV for real-time object detection and classification.

Bounding Box and Labeling:
Implement a bounding box generation mechanism to enclose the objects detected by YOLOv3.

Include a labeling mechanism to classify and assign labels to the detected objects.

Optimization and Efficiency:
Continuously monitor and optimize the system for computational efficiency.

Implement algorithms to dynamically adjust thresholds and parameters based on the real-time performance of the system.

Future Enhancements:
Investigate and integrate feature vectors through Convolutional Neural Networks (CNNs) for more refined motion tracking.

Explore the implementation of linear regression techniques to further enhance the accuracy of motion tracking.

Real-time Car Tracking (Future Work):
Extend the capabilities to become a real-time car tracker by integrating number plate recognition and tracking across multiple CCTV cameras.

Documentation and Reporting:
Maintain comprehensive documentation of the codebase, methodologies, and algorithms used.

Regularly update reports, including findings, challenges, and proposed enhancements.

Collaboration and Iterative Development:
Encourage collaboration among team members for continuous improvement.

Adopt an iterative development approach, incorporating feedback and insights for ongoing enhancements.

3. CONCLUSIONS

This project offers a comprehensive solution for optimizing data storage and improving object detection in traffic camera surveillance. Utilizing YOLOv3, it selectively records relevant data, preventing unnecessary storage. The integration of OpenCV's DNN module enables real-time object detection, accurately identifying and enclosing objects causing changes with precise bounding boxes. The literature review highlights the significance of YOLOv3, COCO dataset, OpenCV, and Neural Networks in enhancing object detection capabilities. The convergence of these frameworks represents a substantial advancement in traffic camera surveillance, addressing storage and detection challenges and contributing to the evolving landscape of computer vision systems.

REFERENCES

[1] D. Kornack and P. Rakic, "Cell Proliferation without Neurogenesis in Adult Primate Neocortex," Science, vol. 294, Dec. 2001, pp. 2127-2130, doi:10.1126/science.1065467.

[2] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[3] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] Redmon, J., & Divvala, S., & Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. arXiv:1506.02640.

[6] Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv:1804.02767.

[7] Salton, G., Wong, A., & Yang, C. S. (1975). A Vector Space Model for Automatic Indexing.

[8] Rao, D., & Shah, M. (2001). View Invariant Representation and Recognition of Actions.

[9] Yilmaz, A., Javed, O., & Shah, M. (2006). Object Tracking: A Survey.