American Journal of Technology Advancement 2997-9382, Vol 03, No. 06, 2026

*Article*

# Algorithms for Automatic Object Recognition in Video Data

**Vohobov Orifjon**
Master's Student in Information Systems, Samarkand State University named after Sharof Rashidov

**Eshtemirov Bunyod Sherali o'g'l**
Doctor of Philosophy (PhD) in Technical Sciences, Research Supervisor.

**Abstract:** This paper analyzes the hardware and software architecture of video data processing and computer vision algorithms. The emergence of video data processing, computer vision algorithms, including deep learning, has revolutionized the field of computer vision, enabling real-time analysis that is essential for traffic management and vehicle identification. This study presents a system that combines vehicle make and model recognition with Automatic Number Plate Recognition (ANPR) and achieves an impressive 97.5% accuracy. Unlike traditional methods, which focus either on make and model recognition or ANPR, this study combines both aspects into a single, coherent system and provides a more comprehensive and efficient solution for vehicle identification, ensuring reliable operation even in harsh weather conditions. The video data processing process consists of several main stages: data acquisition, compression, frame processing, and analysis of the results. The software and hardware used in these processes are of great importance. Hardware components such as GPU (Graphics Processing Units), FPGA (Field Programmable Logic Arrays) significantly accelerate the video data processing process. Software platforms such as TensorFlow, OpenCV, PyTorch also play an important role in the development and implementation of algorithms.[1] The article examines the application of deep learning techniques, including OpenCV and the Python programming language. Thousands of tests were conducted on the system using MobileNet-V2 and YOLOx (You Only Look Once) for vehicle detection, and YOLOv4-tiny, Paddle OCR (Optical Character Recognition), and SVTR-tiny for ANPR

**Keywords:** Video data processing, computer vision, Car model, license plate recognition, deep learning, computer vision, OpenCV, MobileNet-V2, YOLOv4, GradCam. CNN, YOLO, GPU, FPGA, TensorFlow, OpenCV, real-time algorithms, hardware software suite.



This is an open-access article under the CC-BY 4.0 license

## 1. Introduction

Nowadays, as technology is doubling in every aspect of life, services are being further enhanced by using such technologies [1]. Technology also plays a significant role in the road transport system. The road transport system plays a key role in traffic monitoring [2], crime detection systems, stolen vehicle tracking and protection programs. The number of active vehicles has increased dramatically, which has led to more illegal activities. The rapid proliferation of vehicles makes it difficult to track

them, which emphasizes the importance of this for the relevant authorities. As the flow of vehicles increases day by day, the need for automated number plate recognition and vehicle make and model identification is increasing. This tool is also valuable for monitoring vehicle speed, especially in view of the serious accidents that occur every year. The UK police first adopted the Automatic Number Plate Recognition (ANPR) system in 1976, and its deployment has increased significantly in the last decade [1]. Thousands of surveillance cameras are installed along highways, primarily for traffic management and law enforcement. Continuous manual inspection is not feasible because such an approach requires a large amount of effort, which is costly. Automatic visual interpretation offers the possibility of identifying, tracking, and classifying all traffic. A particularly important concept in this context is make and model recognition (MMR). The make and model of a vehicle can be used to identify stolen vehicles by comparing the observed vehicle model with past data. Another application is the identification of specific vehicles after a crime, where only a description of the vehicle is available without the license plate. In such cases, the make and model of the vehicle must be visually obtained [3]. These problems are the main points of this article. Vehicle make and model recognition (VMMR) or vehicle identification solves the problem of identifying the make and model of a vehicle, i.e., identifying the name of the vehicle manufacturer and the specific product, when presented with an image or video of a vehicle. Recently, VMMR has become an increasingly important component of Intelligent Transportation Systems (ITS) and has attracted much attention from the research community [4], [5]. registered data associated with license plate numbers. Another application is the identification of specific vehicles after a crime, where only the description of the vehicle is available without the license plate. In such cases, the make and model of the vehicle must be visually obtained [3]. These problems are the main points of this paper. Vehicle Make and Model Recognition (VMMR) or vehicle identification is the identification of the make and model of the vehicle, i.e.

It solves the problem of identifying the vehicle manufacturer and the exact product when presented with an image or video of a vehicle. Recently, VMMR has increasingly become an important component of Intelligent Transportation Systems (ITS) and has attracted much attention from the research community [4], [5]. Make and model recognition provides important information in many applications such as traffic management, intelligent surveillance, traffic analysis, and traffic monitoring [6]. For example, VMMR technology is used in traffic cameras installed at toll stations to automatically detect and recognize passing vehicles. In traffic monitoring, vehicle model recognition is used to generate vehicle flow statistics that are used by ITS to analyze real-time traffic conditions. In traffic safety [2], make and model recognition helps to find stolen vehicles [7], [8]. Thus, the aim of this paper is to solve the security problem and find cases of missing and mismatched license plates with a precise visual analysis system for stolen vehicles. To this end, we present an MMR system developed for the National Police of Uzbekistan, in which vehicles are tracked from a camera mounted on a highway overhead sign structure to obtain precise make and model information. The obtained data can be combined with existing ANPR data [8]. The main focus of the system implementation is on single-line tracking (see Figure 1). This existing camera is used to provide the training process of our recognition system. The recognition model is trained to recognize vehicles from a large training set of vehicle images, make and model labels. Due to bandwidth constraints between the camera (online) and our training and testing facilities (offline), we need to optimize the collection of training and testing samples. Another challenge is the automated processing of new and unique vehicle models registered in the vehicle registration database, for which it is difficult to collect training and test images. For these reasons, we propose to use datasets for automobiles and ANPR, such as the Coco dataset and the Stanford dataset. In this system

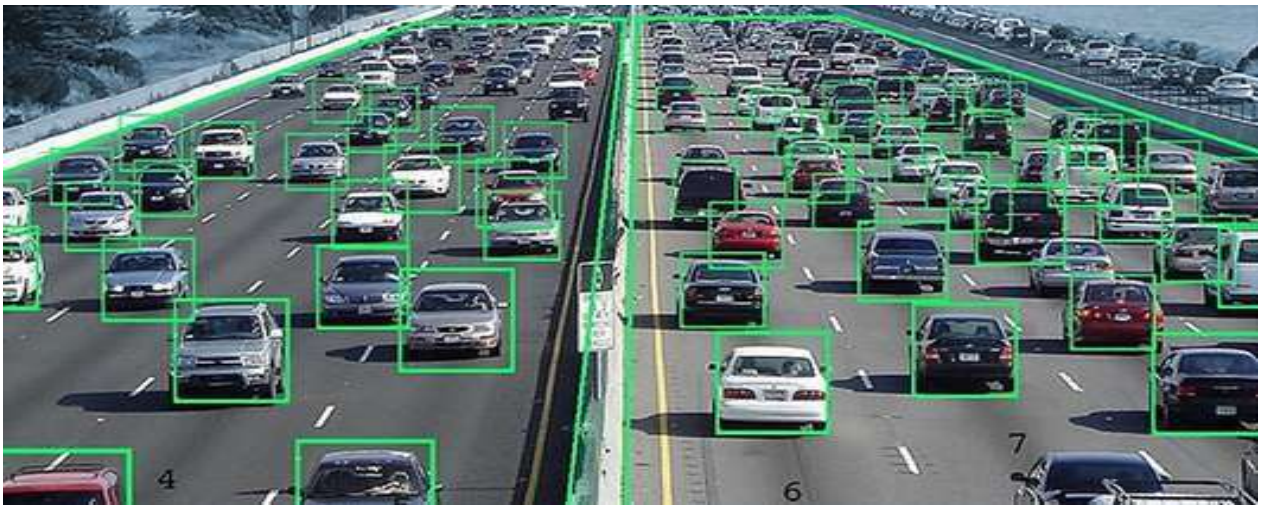



Figure 1 OCR technology used for object recognition [1].

The sampling and their annotations are automated, while the updated training still requires manual control. This approach allows for the creation of an initial dataset and the collection of new vehicle samples over time, ensuring the best performance of the system at all times [9], [10]. In particular, the VMMR system is able to classify vehicles into bounding boxes with a full frontal view.

was developed with a detection and classification stage for placement. The aim of this work is to search for the vehicle brand. Model data that can be implemented in conjunction with the ANPR system. Our two-stage approaches allowed us to detect vehicles for each video frame and perform their classification after the vehicle was found. This paper extends our initial review work by discussing in detail the MMR classification performance and the evaluation of the MMR system. First, we report a comparison between different convolutional neural networks for classifying a vehicle model. Second, we provide more information about the classification performance by finding the most informative region for VMMR classification and measuring the robustness against occlusions. Third, we further investigate the misclassifications in the VMMR classification system to identify shortcomings in the system and information processing [10], [11]. Deep learning is now an effective way to deal with big data, using sophisticated algorithms and artificial neural networks to train machines to learn from experience, classify and recognize images in the same way as the human brain. For deep learning, in particular, the convolutional neural network or CNN is an artificial neural network that is widely used for image/object recognition and classification. Thus, in the context of Deep Learning, objects within an image are mainly recognized using CNN. CNN has found wide application in several tasks such as image processing problems, localization and segmentation in computer vision tasks, video analysis for obstacle detection in self-driving vehicles, speech recognition in natural language processing. Since CNN plays a significant role in rapidly developing and emerging fields, they are often used in Deep Learning [12]. Computer vision is an interdisciplinary field of study that deals with how computers can gain high-level understanding from digital images or videos. Therefore, computer vision tasks include methods for acquiring and processing digital images, analyzing and understanding the various formats they contain, such as single images and video sequences, or extracting information from high-dimensional real-world data to produce numerical or symbolic descriptions that are useful for decision-making. ANPR is a technology that uses optical recognition of characters in images to read vehicle registration numbers. Modern advanced computer vision technology, along with the decline in the cost of related devices in recent years, allows for automatic visual identification of vehicles, whether online or offline, via

video/CCTV. Therefore, this research aims to develop an integrated system that combines vehicle make and model recognition with ANPR, efficient performance and efficiency. with an amazing 97.5% accuracy rate. The system demonstrates excellent robustness in a variety of environmental conditions, including fog, rain, and low light. It has been rigorously tested with thousands of images at the entrance of Firat University, demonstrating its practical applicability. In addition, the use of GradCam technology has been demonstrated in neural network decision-making processes, identifying areas for improvement, Automatic License Plate Recognition (ALPR) - a technology that identifies vehicle license plate numbers from images captured by cameras that can be in color, black-and-white, or infrared formats. This technology uses a number of methodologies, including object detection, image processing, and pattern recognition.

## II. Research Methodology

In the last decade, with the development of information technology and the increase in the population, many methods have been developed and tested to develop systems for recognizing the make, model, and license plate of a vehicle. A vehicle recognition system includes two main stages: detection and classification, which are aimed at identifying and categorizing vehicles from a direct frontal perspective. Various methods are used in the initial detection stage. Ren and Lan [20] use frame difference, while Prokaj and Medioni [21] use background subtraction methods to determine the full dimensions of the vehicle.

**Problem statement:** Using the TensorFlow framework, create a model that classifies videos based on specific categories of actions.

**Problem classification:** Given a sequence of video frames, the model must classify the video to determine what kind of action is taking place in it (for example, walking, running, swimming, etc.). Data generation:

- Y - videos representing a sequence of frames $\{Y_1, Y_2, ..., Y_n\}$, where N is the number of frames (A × B × $C$ pixel matrix, where A, B are the height and width of the frame, and C is the number of channels).
- Z is a symbol indicating the category of movement.

**Mathematical model.**

Data included:

$$Y = \{Y_1, Y_2, ..., Y_n\} \quad (1)$$

Character separation:

- A CNN convolutional neural network is used to separate the characters in the frame $(Y_1, Y_2, ..., Y_n)$ $f(y_t\theta)$ this $\theta$ with parameters A function representing the convolution and operations of a frame. Pooling of time-dependent symbols:
- A recurrent neural network (RNN) such as LSTM (Long Short-Term Memory) is used to process time-dependent data:

$$\hat{y} = softma(W \cdot a_t + k) \quad (2)$$

**1. Classification:**

The last hidden state of $a_t$ is used for class prediction:

$$\hat{y} = softma(W \cdot a_t + k) \quad (3)$$

where W and k are the parameters of the linear layer.

2. **Loss function:**

$$L = \frac{1}{n} \sum_{i=1}^{n} \sum_{J=1}^{K} yij \log\left( \hat{yij} \right) \quad (4)$$

where n is the number of video sequences in the learning competition, k is the number of classes.

**Algorithmic approach:**

**YOLO algorithm:**

The YOLO model is designed for real-time object detection in video frames, and finds objects based on a single analysis. This algorithm is characterized by high speed and accuracy, which is why it is effective for real-time processing processes.

Below is a block diagram of the algorithmic approach developed on the basis of the TensorFlow platform (Figure 1) Convolutional Neural Networks (CNN): The CNN algorithm was used in the process of object detection and classification in video streams. This algorithm processes data through convolution, pooling, and classification stages, extracting important features of objects.

Speed and processing time: The video stream processing speed was measured in FPS (Frames per second). The model's processing speed and real-time performance were analyzed. Experimental methodology: - Video stream images were analyzed in real-time and algorithms (CNN, YOLO) were trained and optimized. - Experiments on object detection and segmentation in video streams were conducted using GPU and FPGA devices. - The results of each experiment were evaluated for accuracy, speed, and efficiency and analyzed in practical applications. The results showed that using GPU and FPGA hardware, video data processing algorithms were significantly accelerated and accuracy increased. Also, the correct selection of software platforms allows for the effective use of algorithms

**Results and Discussion** This study analyzed the effectiveness of video data processing and the use of computer vision algorithms, as well as the integration of hardware and software tools. The results and discussion are as follows: Efficiency of GPU and FPGA technologies: During the experiment, the efficiency of GPU and FPGA technologies in the process of processing video data was analyzed. High speed and accuracy were observed when running convolutional neural networks (CNN) and YOLO algorithms using GPU. Using NVIDIA RTX 3090 GPU, the FPS (frames per second) performance of the model increased significantly and high accuracy was achieved when processing video streams. Real-time performance of up to 30 FPS was recorded using GPU hardware, which ensured smooth processing of video streams

YOLO (You Only Look Once): The YOLOv5 model has shown its effectiveness in analyzing video streams in real time. In experiments conducted using YOLOv5, the object detection rate was high, with an mAP index of 76%. This algorithm has shown to be especially effective in detecting multiple objects in video frames at the same time. The YOLO algorithm can be successfully applied in video surveillance systems and security.[8] Software platforms: TensorFlow and Keras: In experiments conducted using these platforms, significant progress has been made in creating and training neural networks. Using TensorFlow and Keras, the process of developing CNN networks has been simplified and model optimization has become more efficient. The optimization of the training process has helped to increase the speed and accuracy of the model. OpenCV: Pre-processing of video streams was performed using OpenCV. OpenCV was very effective in separating frames, applying

filters, and improving image quality. This library expanded the ability to work in real time and made the software easier to use. Practical application: The practical application of video data processing and computer vision algorithms was analyzed.

**References:**

[1] Speech-to-Text Models in Uzbek Language: Achievements and Limitations F Nazarov, A Soliev, B Eshtemirov
[2] 2026 International Russian Smart Industry Conference (SmartIndustryCon Ahmadov, A. (2020). "Fundamentals of Artificial Intelligence and Its Applications." Tashkent: Ma'naviyat.
[3] Alimov, N. (2019). "Development and Applications of Computer Vision Systems." Journal of Innovative Technologies, 4(3), 45- 52.
[4] Detection and analysis of traffic jams using computer vision technologies A Akhatov, F Nazarov, B Eshtemirov
[5] Artificial Intelligence, Blockchain, Computing and Security Volume 2, 761-765
[6] Bektemirov, I. (2021). "The Role of GPU and FPGA Technologies in Real-Time Processing." Publishing House of the Academy of Sciences of Uzbekistan.
[7] Gulomov, U. (2018). "Application of Neural Networks and Computer Vision Algorithms in Industry." Journal of Technological Processes and Innovations, 6(2), 55-63. 5
[8] Karimov, M. (2020). "Theoretical Foundations of Computer Vision Technologies and Algorithms." Samarkand: Ilm va Fan.
[9] Akhatov, F.M. Nazarov and B. Eshtemirov, "Traffic jam detection and analysis using computer vision."
[10] L. Ramachandran, S. Sukumaran and S.R. Sunny. Intersection-based traffic-aware routing with low overhead in VANET. International Journal of Digital Information and Wireless Communications (IJDIWC), Vol. 3, No. 2, pp. 190–196, 2013.
[11] Guo, J. Huang, W. Williams, B.M. An adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification. Transp. Res. Part C 2014, 43, 50–64.
[12] Wu, B.; Nevatia, R. Detecting multiple, partially occluded people in a single image using a combination of Bayesian edge detectors. In Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV 2005), Beijing, China, 17–21 October 2005; pp. 90–97. [7] Raj, Jithin, HareeshBahuleyan, and Lelitha Devi Vanajakshi. "Application of data mining techniques for traffic density estimation and prediction." Transportation Research Proceedings 17 (2016): 321–330.
[13] DETERMINING THE LEVEL OF ENVIRONMENTAL POLLUTION USING BIG DATA-BASED ANALYSIS B Eshtemirov
[14] Ramesh, K., Bhimisetty Anil Kumar, and P. N. Renjith. "Review-based protection against password guessing attacks and phishing attacks via various online records." 2020 International Conference on Inventive Computation Technologies (ICICT), pp. 824–827. IEEE, 2020.
[15] Guowen Dai, Changxi Ma, and Xuecai Xu. A method for predicting short-term traffic flow based on spatiotemporal analysis and grouping for urban roads. IEEE Access, 7:143025–143035, 2019.
[16] Weibin Zhang, Yinghao Yu, Yong Qi, Feng Shu, and Yinhai Wang. Short-term traffic flow prediction based on spatiotemporal analysis and deep learning. Transportmetrica A: Transport Scienc,, 15(2):1688–1711, 2019.
[17] Stages and algorithms for estimating traffic jams on roads using artificial intelligence: Stages of estimating traffic jams on roads using artificial intelligence …E Bunyod, A Akhatov
[18] MODERN PROBLEMS AND PROSPECTS OF APPLIED MATHEMATICS 1 (01)
[19] Liu, Yunxiang, and Hao Wu. " Traffic congestion prediction based on random forest." 2017 10th International Symposium on Computational Intelligence and Design (ISCID). Vol. 2. IEEE, 2017.
[20] Xiong, Y. Research on predicting e-coupon usage based on XGBoost. Computer Science and Applications, 9(5), 1029–1035. DOI 10.12677/CSA.2019.95116.