Deep Learning for Traffic Congestion Detection: A Survey Paper

Sternford Mavuchi[1], Tirivangani Magadza[2], Racheal Chikoore[3]
[1]Department of Software Engineering, [2,3]Department of Information Technology,
School of Information Science And Technology, Harare Institute of Technology

Abstract:- Traffic congestion is a major problem in urban areas, leading to increased travel time, economic losses, and environmental pollution. By analyzing traffic data from traffic cameras, we can detect and predict traffic congestion with high accuracy. In this survey, we explore the use of deep learning techniques for traffic congestion detection. Deep learning models, such as convolutional neural networks and recurrent neural networks, have shown promising results in traffic congestion detection. We also discuss the challenges and future directions of this field, including the need for high-quality data and the development of real-time traffic management systems.

Keywords:- *Traffic Congestion, Deep Learning, Machine Learning, Artificial Neural Networks, Computer Vision, Traffic Cameras, Traffic Data, Traffic Management, Real-Time Systems, Convolutional Neural Networks, Recurrent Neural Networks, Data Analysis, Pattern Recognition, And Image Processing.*

I. INTRODUCTION

Traffic congestion is a situation when the demand for road space exceeds supply [1],and is reflected by an excess of vehicles on a portion of roadway at a particular time resulting in speeds that are slower or sometimes slower than normal. In some context, traffic congestion refers to the state of heavy traffic flow on urban roads and highways, which can cause delays, accidents, and environmental pollution[2]. With the progression of time, there has been a huge increase in traffic in different parts of Zimbabwe specifically in Harare leading to high cases of traffic congestion. The effects of traffic congestions resulted in massive delay on the roads leading to increased travel time, economic losses and increase in environmental pollution[1]. Among the causes of traffic congestion, there exist lack of adequate infrastructure such as traffic lights, lane markings, road signs, unavailability of electricity to power traffic lights and poor road rehabilitation.

The implementation of traffic management strategies is of great urgency to mitigate the side effects of traffic congestion. The government have put measures in place to deal with traffic congestion such as the introduction of a taskforce within the Zimbabwe Republic to decongest the Harare CBD, the joint taskforce between the Zimbabwe republic police traffic department with the Harare City Council traffic police to manage congestion at intersections. The ongoing roadway rehabilitation, implementation of detours, traffic interchange and many more. But it is difficult to plan and propose solutions related to traffic congestion if there is no documented quantities of roadway and intersection congestion.

However, this study is aimed at improving the already existing measures in trying to detect and manage traffic congestion in real time. This project will address how deep learning a field in artificial intelligence can be employed to detect traffic congestion and provide actionable information to the responsible authorities to act and curb building congestion or already build up congestion among other means of managing and controlling it.

II. LITERATURE REVIEW

A. *Traffic congestion as a concept*

Congestion can be described as a condition that occurs when the output capacity of a facility becomes less than the input capacity[1], [3]. The literature on congestion detection by videos has adopted a common pipeline in estimating traffic congestion[2]. This is through the extraction of the visual features to estimate the traffic state e.g., traffic speed, density, flow or in terms of derived variables such as the ratio of average speed to speed limit[3]. An overview of the traffic parameters used to assess congestion is provided in article [4]. To define a classification task, once a variable has been selected, the values of the variable are quantified into a predetermined number of levels. For example, a traffic density of more than a certain threshold can be referred to as congested or not congested. In exploring the literature further, traffic congestion was classified into two types, recurring and nonrecurring[1], [4], [3]. Recurring congestion occurs frequently. This type of congestion generally congestion at the same place and time every day, and it is generally the consequence of factors that act regularly or periodically on the transportation system such as daily commuting or weekend trips.On the contrary, on-recurring congestion is the result of unexpected or unplanned significant events, such as road construction, accidents, special occasions, and other associated events or activities that have an unpredictable and erratic impact on the transportation system.[1] and[5]presents a summary of the factors that contribute to both types of congestion. They report that recurring congestion is usually caused by an infrastructural constraint and the non-recurring congestion can be caused by unplanned occurrences like severeweather, man-made or natural disasters, accidents, or planned activities like large concerts and road rehabilitation works.

B. Methods of traffic congestion detection

In developed countries, CCTV infrastructures are being utilised to aid the vision based traffic detection of traffic parameters such as volume, density and speed[6], [7] . Preliminary investigations saw deep-learning as a promising technique for traffic congestion detection [6].

In some literature, the traditional methods that were being used to estimate traffic state were point-based sensors which including inductive loops, piezoelectric sensors and magnetic loops. With advancement in active infra-red or laser radar sensors have led to the gradual replacement of the traditional point-based sensors. Also, with the increasing usage of navigation-based GPS devices, probe-based data are emerging as a cost-effective way to collect network-wide traffic data [6].

➢ *Detection based methods*

Detection-based methods use individual video frame to identify and localize vehicles and thereby perform a counting task [6]. In[8], they try to identify and localize vehicles in each frame. Ozkurt et al used neural network (Faster RCNN) methods to perform vehicle counting and classification tasks from video records [9]. Kalman filter-based background estimation has also been used to estimate vehicle density [10] . However, they were found to perform poorly for videos with low resolution and high occlusion [7].

➢ *Motion based methods*

Quite a number of motion-based methods have been put forward to estimate traffic flow utilizing vehicle tracking information. Asmaa et al used microscopic parameters extracted using motion detection in a video sequence [6], [11]. They also analyzed the global motion in the video scene to extract the macroscopic parameters. However, these methods tend to fail due to lack of motion information and low frame rates of videos, some vehicles appear only once in a video, and hence, it becomes difficult to estimate their trajectories[7].

- **Holistic approaches:** These techniques perform analysis on the whole image, thereby avoiding segmentation of each object. uses a dynamic texture model based on Spatiotemporal Gabor Filters for classifying traffic videos into different congestion types, but it does not provide accurate quantitative vehicle densities. formulates the object density as a linear transformation of each pixel feature, with a uniform weight over the whole image. It suffers from low accuracy when the camera has large perspective [6], [7].

- **Deep learning techniques:** With increasing access to new data sources, new opportunities to automatically detect traffic congestion are being explored. The most commonly used data sources to detect both types, recurring and non-recurring congestion are images and videos obtained from traffic cameras [3].

In this study[2], deep learning is defined as type of machine learning that uses artificial neural networks to learn and recognize patterns in data. The authors use deep learning techniques to identify vehicles and classify traffic states based on visual features extracted from surveillance camera images.

Deep learning models which are known to perform well for Computer Vision tasks have been adopted to detect traffic congestion[3]. Computer vision refers to the task of extracting useful information from images.

After careful analysis of literature, we found out that Convolutional Neural Networks forms the building block of commonly used deep learning architectures for image classification tasks.

In [12], the author explained a brief background to the Convolutional Neural Networks (ConvNets or CNNs).

The three primary layers of a common ConvNet architecture are convolution, max-pooling, and fully connected layers. The fundamental building block of ConvNets are convolution layers. They are employed in the process of filter-based picture feature extraction.

Convolutional layers generate activation or feature maps, which are then sampled down using pooling layers. While strides (more than 1) in a conventional convolution layer can also be used to achieve down sampling, max-pooling layers introduce translational invariance that improves model generalisation at the expense of spatial inductive bias and have no learnable parameters. For the goal of classification, fully connected layers are employed (connecting learned characteristics with their appropriate labels). Usually, the softmax activation function is used to activate the final completely connected layer in classification settings.

The author provided an example, saying that AlexNet and VGG are two ConvNets architectures that adhere to the aforementioned framework. Most contemporary ConvNet architecture are more complex than a simple stack of fully-connected layers, max-pooling, and convolution. Architectures such as ResNet and similar networks, for instance, use residual connections [12].

The foundational studies in this area were AlexNet [13], Resnet [14], VGGNet [15] and YOLO [16]. These models mostly pre-trained on large image datasets like ImageNet and COCO and are readily available.

In light of the modern ConvNet architectures, author in [6] adopted the YOLO model for general purpose congestion detection and localization from CCTV video feeds. Current object detection systems repurpose powerful CNN classifiers to perform detection. For example, to detect an object, these systems take a classifier for that object and evaluate it at various locations and scales in the test image. YOLO reframes object detection; instead of looking at a single image 1000 times to accomplish detection, it looks at an image only once to perform the full detection pipeline. A single convolutional network simultaneously predicts multiple bounding boxes and class probabilities for those boxes. This makes YOLO extremely fast and easy to generalize to difference scenes. In another context, [17] they carried out the detection process using

YOLOv3 which requires adequate computing devices. TOLOv3 is a method that uses deep learning to recognize objects. Tests were carried out by utilizing the CPU and GPU variations of the Jetson Nano.

Furthermore, [2]compared the performance of different deep learning models, such as You Only Look Once (YOLO), the Single Shot MultiBox Detector (SSD), and Mask R-Convolutional Neural Networks (R-CNN), and find that they outperform classic machine learning approaches in terms of accuracy. The deep learning method achieves an accuracy of 99.9% for binary traffic state classification and 98.6% for multiclass classification.

When combined, DCNNs represent an advanced method for classifying images and detecting objects [6]. They employed a conventional ConvNet architecture with convolution and pooling layers in [6]. The input photos in this investigation varied in size because multiple cameras' images were used. In another piece of literature [2], the Single Shot MultiBox Detector (SSD) was built using a Convolutional Neural Network (CNN) as its foundation. This produced a set of fixed-size bounds, and a confidence score that indicated the likelihood that the object inside the box belonged to a particular class was given. To forecast things on several scales, the CNN is composed of multiple layers that gradually get smaller. Vehicle detection using SSD has been done successfully in [18].

Furthermore, in [19], two variants of CNN-based architectures (AlexNet and YOLO) are used to identify congestion by applying a binary classification to traffic photos gathered over a six-month period from 121 cameras in Iowa, USA. It takes a lot of time to manually identify traffic photos as either congested or not. The authors consequently automatically classify the photos into two groups according to occupancy (occupancy is marked as "congested") using occupancy data collected using vehicle loop detectors (VLDs). For AlexNet and YOLO, respectively, the stated accuracy for identifying congestion was 90.5% and 91.2%.

In order to identify traffic congestion on traffic photos collected from more than 100 cameras in the Chinese province of Shaanxi, [18]compare two versions of AlexNet and VGGNet.Their dataset is incredibly diverse, with photos showing both day and nighttime traffic as well as a range of weather situations. According to their findings, both architectures perform similarly (78% for AlexNet and 81% for VGGNet). They claim that because AlexNet’s neural network is smaller, it trains far more quickly. They categorise things using binary terms ("jam" or "no jam").

Researchers in[2]evaluated the effectiveness of Mask-RCNN and YOLO on three datasets that were manually tagged and came from two different sources of traffic image data (GRAMME and Trafficdb). The three datasets have different picture quality: the first has 23435 low resolution images (480 x 320 pixels), the second has 7520 mid resolution images (640 x 480 pixels), and the third has 9390 high resolution images (1280 x 720 pixels). They take two stages to detect congestion.Determining the total number of vehicles in every frame is the first step. In this stage, YOLO achieves an accuracy of 82%, 86%, and 91%, whereas the Mask-RCNN achieves 46%, 89%, and 91%, respectively. The training time of YOLO was nearly half that of Mask-RCNN, and its performance was independent of image quality. They employ YOLO as the object detector model and feed the second step's output into it. The second phase used Resnet on the YOLO output to predict traffic congestion as a three-class multiclass classification task. For low, medium, and heavy congestion, the stated accuracy is 99.7%, 97.2%, and 95.9%, respectively.

[19] classified traffic photos taken in Jakarta, Indonesia, using a CNN model. Fourteen camera locations were used for the fifteen days that the data were collected. The traffic photos were manually labelled into "jammed" and "not jammed" classes. 89.5% is the stated average accuracy while utilising 10-fold cross-validation.

In 2020, [20]looked into how well YOLO performs in situations with a lot of variation in traffic. They gather information from Karnataka, India, for a week. Using a YOLO model that has been pre-trained on the COCO dataset, they employ transfer learning. When it comes to identifying the different types of vehicles in each zone, YOLO does a good job of counting cars, trucks, and buses in the photographs (accuracy ranging from 92% to 99%).

Table 1: Summary of papers on Deep learning for traffic congestion detection

Paper	DNN architecture	Data Source	Performance	Evaluation Metrics
[2]	YOLO, R-CNN	TrafficDB, GRAM RTM	YOLO 99.9%,	Jaccard Index, Intersetion over Union (IoU)
[19]	CNN	Traffic camera images	CNN 89.50%	Stratified k-fold cross-validation
[20]	YOLO	COCO	YOLO 99%	Accuracy
[21]	YOLO, DCNN	Traffic Data	YOLO 91.5, DCNN 90.5	Precision, Recall, Accuracy
[22]	AlexNet, ResNet, InceptionNet, R-CNN, Mask-RCNN, VGGNet, YOLO	ImageNet COCO, Traffic camera images	AlexNet 92.5%, ResNet 95.5%, InceptionNet 94.5%, R-CNN 92.5%, Mask-RCNN 95.5%, VGGNet 93.5%, YOLO 94.5%	True Positive Rate (TPR), True Negative Rate (TNR), Accuracy
[23]	SSD	Traffic pictures		mAP

III. CHALLENGES AND FUTURE DISCUSSIONS

We notice significant differences in the image quality depending on the data source. As a result, the model performs differently. One major challenge is the notable diversity of the traffic flow in traffic images from underdeveloped countries. While deep learning-based high-resolution has been thoroughly investigated in computer vision, its use to improve the quality of traffic images is still in the early stages. To enhance the model's performance, further research must be done to enhance the quality of the images or videos.

In addition, some of the deep-learning architectures require huge computational resource such as GPU to train on due to huge datasets with high resolution videos or images, number of epochs etc.

IV. SUMMARY

In this survey, we have explored the use of deep learning techniques for traffic congestion detection. Our analysis has shown that deep learning models, such as convolutional neural networks and recurrent neural networks, can accurately detect and predict traffic congestion using traffic data from various sources, including traffic cameras.The use of YOLO architecture models has gained much traction according to the number of papers reviewed. This is because of its performance in different experiments. We have also discussed the challenges and future directions of this field, including the need for high-quality data and the development of real-time traffic management systems. Overally, our findings suggest that deep learning has the potential to revolutionize traffic management and improve the quality of life in urban areas. By implementing these techniques, we can reduce travel time, economic losses, and environmental pollution, leading to a more sustainable and efficient transportation system.

REFERENCES

[1]. S. Munuhwa, K. Muchenje, J. Pule, T. Mandere, and T. Gabakaiwe, “Approaches for Reducing Urban Traffic Congestion in the City of Harare,” Feb. 2020, doi: 10.7176/JESD/11-4-01.

[2]. D. Impedovo, F. Balducci, V. Dentamaro, and G. Pirlo, “Vehicular Traffic Congestion Classification by Visual Features and Deep Learning Approaches: A Comparison,” *Sensors*, vol. 19, no. 23, p. 5213, Nov. 2019, doi: 10.3390/s19235213.

[3]. N. Kumar and M. Raubal, “Applications of deep learning in congestion detection, prediction and alleviation: A survey,” *Transp. Res. Part C Emerg. Technol.*, vol. 133, p. 103432, Dec. 2021, doi: 10.1016/j.trc.2021.103432.

[4]. T. Afrin and N. Yodo, “A Survey of Road Traffic Congestion Measures towards a Sustainable and Resilient Transportation System,” *Sustainability*, vol. 12, no. 11, Art. no. 11, Jan. 2020, doi: 10.3390/su12114660.

[5]. J. McGroarty and N. Urban, “Recurring and Non-Recurring Congestion,” 2010. Accessed: Oct. 12, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Recurring-and-Non-Recurring-Congestion-McGroarty-Urban/817f5f8aa6e3eb9947e56b51919714b1daf42d9b

[6]. P. Chakraborty, Y. Okyere, S. Poddar, V. Ahsani, A. Sharma, and Chakraborty, “Traffic Congestion Detection from Camera Images using Deep Convolution Neural Networks,” *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2672, Jun. 2018, doi: 10.1177/0361198118777631.

[7]. S. Zhang, G. Wu, J. P. Costeira, and J. M. F. Moura, “Understanding Traffic Density From Large-Scale Web Camera Data,” presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 5898–5907. Accessed: Feb. 01,

2023. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2017/html/Zhang_Understanding_Traffic_Density_CVPR_2017_paper.html

[8]. Y. Zheng and S. Peng, "Model based vehicle localization for urban traffic surveillance using image gradient based matching," in *2012 15th International IEEE Conference on Intelligent Transportation Systems*, Sep. 2012, pp. 945–950. doi: 10.1109/ITSC.2012.6338660.

[9]. C. Ozkurt and F. Camci, "Automatic Traffic Density Estimation and Vehicle Classification for Traffic Surveillance Systems Using Neural Networks," *Math. Comput. Appl.*, vol. 14, no. 3, Art. no. 3, Dec. 2009, doi: 10.3390/mca14030187.

[10]. M. Balcilar and A. C. Sonmez, "Extracting vehicle density from background estimation using Kalman filter," in *2008 23rd International Symposium on Computer and Information Sciences*, Oct. 2008, pp. 1–5. doi: 10.1109/ISCIS.2008.4717950.

[11]. "Road traffic density estimation using microscopic and macroscopic parameters," *Image Vis. Comput.*, vol. 31, no. 11, pp. 887–894, Nov. 2013, doi: 10.1016/j.imavis.2013.09.006.

[12]. J. de D. Nyandwi, "Modern Convolutional Neural Network Architectures." Sep. 07, 2023. Accessed: Sep. 08, 2023. [Online]. Available: https://github.com/Nyandwi/ModernConvNets

[13]. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2012. Accessed: Oct. 20, 2023. [Online]. Available: https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

[14]. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition." arXiv, Dec. 10, 2015. doi: 10.48550/arXiv.1512.03385.

[15]. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition." arXiv, Apr. 10, 2015. doi: 10.48550/arXiv.1409.1556.

[16]. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection." arXiv, May 09, 2016. doi: 10.48550/arXiv.1506.02640.

[17]. R. G. Putra, W. Pribadi, I. Yuwono, D. E. J. Sudirman, and B. Winarno, "Adaptive Traffic Light Controller Based on Congestion Detection Using Computer Vision," *J. Phys. Conf. Ser.*, vol. 1845, no. 1, p. 012047, Mar. 2021, doi: 10.1088/1742-6596/1845/1/012047.

[18]. Q. Wang, J. Wan, and Y. Yuan, "Locality constraint distance metric learning for traffic congestion detection," *Pattern Recognit.*, vol. 75, pp. 272–281, Mar. 2018, doi: 10.1016/j.patcog.2017.03.030.

[19]. J. Kurniawan, S. Syahra, C. Kusuma, and A. Afiahayati, "Traffic Congestion Detection: Learning from CCTV Monitoring Images using Convolutional Neural Network," *Procedia Comput. Sci.*, vol. 144, pp. 291–297, Jan. 2018, doi: 10.1016/j.procs.2018.10.530.

[20]. R. C R and C. Shantala, "Vehicle Density Analysis and Classification using YOLOv3 for Smart Cities," Nov. 2020, pp. 980–986. doi: 10.1109/ICECA49313.2020.9297561.

[21]. P. Chakraborty, Y. Okyere, S. Poddar, V. Ahsani, A. Sharma, and Chakraborty, "Traffic Congestion Detection from Camera Images using Deep Convolution Neural Networks," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2672, Jun. 2018, doi: 10.1177/0361198118777631.

[22]. N. Kumar and M. Raubal, "Applications of deep learning in congestion detection, prediction and alleviation: A survey," *Transp. Res. Part C Emerg. Technol.*, vol. 133, p. 103432, Dec. 2021, doi: 10.1016/j.trc.2021.103432.

[23]. Q. WU and S. LIAO, "Single Shot MultiBox Detector for Vehicles and Pedestrians Detection and Classification," *DEStech Trans. Eng. Technol. Res.*, Feb. 2018, doi: 10.12783/dtetr/apop2017/18705.