International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

A DEEP LEARNING SYSTEM FOR VEHICLE SPEED DETECTION VIA OBJECT RECOGNITION AND DEPTH ESTIMATION

[1]Dr. M. Sridhar, [2]Mr. M. Rafath Kumar, [3]M.V.Satya Naga Sai Ganesh,[4]Mekapotula Preetham, [5]Peddagoundla Sai Shiva Kumar,

[1] Professor, Rajamahendri Institute of Engineering & Technology, Bhoopalapatnam, Near Pidimgoyyi,Rajahmundry,E.G.Dist.A.P. 533107.

[2]Assistant Professor, Dept. of CSE, Rajamahendri Institute of Engineering & Technology, Bhoopalapatnam, Near Pidimgoyyi, Rajahmundry, E.G. Dist. A.P. 533107.

[3,4,5] Students, Dept. of CSE, Rajamahendri Institute of Engineering & Technology, Bhoopalapatnam, Near Pidimgoyyi, Rajahmundry, E.G. Dist. A.P. 533107.

Abstract—
Road accidents are quite common in almost every part of the world, and, in the majority, fatal accidents are attributed to over speeding of vehicles. The tendency to over speeding is usually tried to be controlled using check points at various parts of the road but not all traffic police have the device to check speed with existing speed estimating devices such as LIDAR based, or Radar based guns. The current project tries to address the issue of vehicle speed estimation with handheld devices such as mobile phones or wearable cameras with network connection to estimate the speed using deep learning frameworks

I. INTRODUCTION

As with any modern transportation system, maintaining safe driving conditions is of paramount importance [1]. Ensuring compliance with speed restrictions, reducing accidents, and improving traffic flow requires the capacity to correctly detect and monitor vehicle speeds on roads and highways. The need of efficient speed enforcement tactics is underscored by studies that indicate speeding is a primary cause of road fatalities and injuries. Accidents can be prevented and their effects lessened with the use of speed detection systems, which discourage speeding and promote compliance with speed restrictions. Here is where a reliable mechanism for tracking speed becomes crucial.
Radar, light-based imaging and detection radar (LIDAR), and video-based systems are some of the technologies used by modern vehicle speed detection systems. To determine a vehicle's speed, radar systems detect the Doppler change in the frequency of reflected signals, which are radio waves. Lidar systems use laser beams to measure how fast a vehicle is moving by timing how long it takes for light pulses to bounce off the object and back to the sensor. To detect and gauge a vehicle's velocity, video-based systems employ a combination of cameras and image processing algorithms. The third group includes video-based systems, which is where our present research fits in. The suggested project incorporates state-of-the-art technologies like OpenCV, YoLOv8, and Convolutional Neural Networks (CNN) to conquer the speed detection

difficulty. In order to prepare films for additional analysis, the project makes good use of OpenCV to extract individual frames. To further enhance the accuracy of speed detection, YOLOv8, which is well-known for its object identification skills, is used to correctly identify automobiles in the video clip. Using depth estimation [4] to more accurately predict the vehicle's speed is also part of the research.
Even though deep learning-based approaches with a fixed camera and position were offered, the researchers are unaware of any prior work of this kind. By removing the limitations of a stationary camera and location, this work aimed to develop deep learning-based models that could detect speed using commonly available handheld devices, such as smartphones. Here is how the remainder of the paper is structured. There was a passing reference to previous research in section II. The technique and experimental findings have been thoroughly explained in Section III and Section IV, respectively. portion V, the last portion, presented the last thoughts.

II. RELATED WORK

The idea of detecting the speed of a vehicle is not novel. A number of techniques exist for determining an object's velocity in motion, including the LIDAR gun [5], the radar gun [6], and the manual counting approach. On the other hand, these approaches are either more expensive or area-specific, meaning that speed can only be predicted at given locations. Proposed approaches for video-based speed estimation are cost-effective and, when trained with a big enough dataset, can achieve high levels of accuracy. Traditional approaches mostly relied on tracking moving vehicles [7] and calculating how long it would take for a vehicle to travel a certain distance [8]. The idea of object detection utilizing different pre-trained models (e.g., YOLOv5, YOLOv8, SSD, Faster RCNN, etc.) was fundamental to all of these approaches. However, the identical location-specific limitation persisted across all of these approaches, necessitating the installation of a camera at a predetermined height. So, it was determined that traffic cops could use their cell phones or any camera with an internet connection to make educated guesses about how fast cars were going. Therefore, it was believed that depth estimates would be relevant in these scenarios in addition to object identification. Due to the ongoing maintenance of the MiDAS [9] python package, it was chosen for depth estimate after extensive study of related research.

III. DATA AND METHODOLOGY

The research field places a premium on high-quality data. The data used in this study came from the main source of data collecting. By meticulously selecting films of moving automobiles, our study was able to gather around 100 data points. Mobile phones with a 640 x 480 pixel resolution were used to record the videos. The movies used for this investigation were carefully selected to ensure that just one automobile was visible in each frame. The cameras in each video were angled either frontally or laterally. From three to seven seconds long, the videos ranged in length. Accompanying the footage of the moving car, the data collecting method also captured the vehicles' speeds.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

The current research relies on data obtained in a manner that ensures each video only contains one automobile. Since vehicle tracking would be crucial in a multi-vehicle scenario, it was believed that this concept could not be used to such a situation. After gathering the selected films, they were sent via the OpenCV framework that included the YOLOv8 object identification model. The YOLOv8 was designed to generate a more condensed bounding box surrounding the vehicle. The method did reveal, however, that in certain movies even the rearview mirrors were being mistaken for objects. As a result, we required threshold confidence to keep just the primary vehicle. In order to determine the threshold confidence probability provided by the YOLOv8 model, manual checks were carried out. We determined that 0.7 was the optimal threshold confidence; that is, we should only keep items for which this value was more than 0.7. Based on the idea that the bounding box should grow in size as the car gets closer to the camera, hit a maximum value, and then shrink in size as the car moves away from the camera, the speed estimation process could proceed. The change in bounding box area is greater for a car driving quickly than for a sluggish automobile going slowly over the same amount of time. On top of that, the average distance between the automobile's pixels and the camera should decrease when the car approaches the camera at a specific speed. The speed of the vehicle should determine this reduction. Therefore, with 't' standing for time in seconds, □□ for the change in size of the bounding box, and □□ for the change in average distance of the automobile from the camera, we may describe speed as a linear function, as indicated in Equation 1.

$$Speed = \beta_0 + \beta_1 t + \beta_2 \Delta A + \beta_3 \Delta D + \epsilon \qquad (1)$$

This is where the mistake is. Since the films were cut for various durations and the frames per second were known, calculating □□ was a breeze. Thus, the value of 耆□ was determined by subtracting the car's bounding boxes from the beginning and end frames of the movie, respectively. Nevertheless, it was quite difficult to determine □□. Each pixel's distance from the camera may be determined via depth estimation. Redshift is applied to pixels that are closer to the camera, and blueshift is applied to pixels that are farther away. The backdrop pictures within the bounding box were obstructing the vehicle's distance calculation from the camera. Instead of using bounding boxes, we looked at the vehicle's mask with the MiDAS output to tackle this problem more thoroughly. We utilized Mask-RCNN to obtain the vehicle mask. Hide face While RCNN and YOLOv8 both use object detection in images, RCNN uses object masking to identify objects rather than bounding boxes. In several object recognition settings, the mask's attempt to produce an enclosed object contour is useful. We used the beginning and ending frames of the films with bounding boxes and vehicle masks to estimate the vehicle's starting and final average distance. Figure 1 shows the whole procedure. We utilized three object detection models—YOLOv8, YOLOv5, and SSD—to conduct a comparison analysis. In order to compare the performances of the models, □pp□□ and □□□□ were taken into account. In order to assess the recommended models' predictive potential, this study took into account both linear and polynomial regression. By retaining the maximum order at 3, we were able to restrict the number of features in polynomial regression.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

IV. EXPERIMENTAL RESULTS

As said before, a total of 90 data points were gathered for this investigation. We did not use advanced machine learning models since there just weren't enough data points for them to avoid overfitting and find the pattern. After testing, we discovered that the YOLOv8 model gave a more precise bounding box around the cars. Since YOLOv8 also returned the coordinates of the bounding box, calculating its area was a breeze. Figures 2 and 3 display the results of YOLOv8 for two of these vehicles. Vehicle footage was gathered from two angles—the front and the side—for this research. Depth estimate was the subject of the subsequent section. To do this, we computed the distances of each pixel from the camera using the MiDAS program. In Figures 4 and 5, you can see the two automobiles' distance maps. Please note that the red areas indicate pixels that are closer to the camera, while the blue areas denote pixels that are further away from the camera. It should be mentioned that only pre-trained models were utilized in this investigation because there wasn't enough data. Focusing just on the objects of interest and minimizing surrounding images was crucial, since the depth map clearly showed that they were highly apparent.

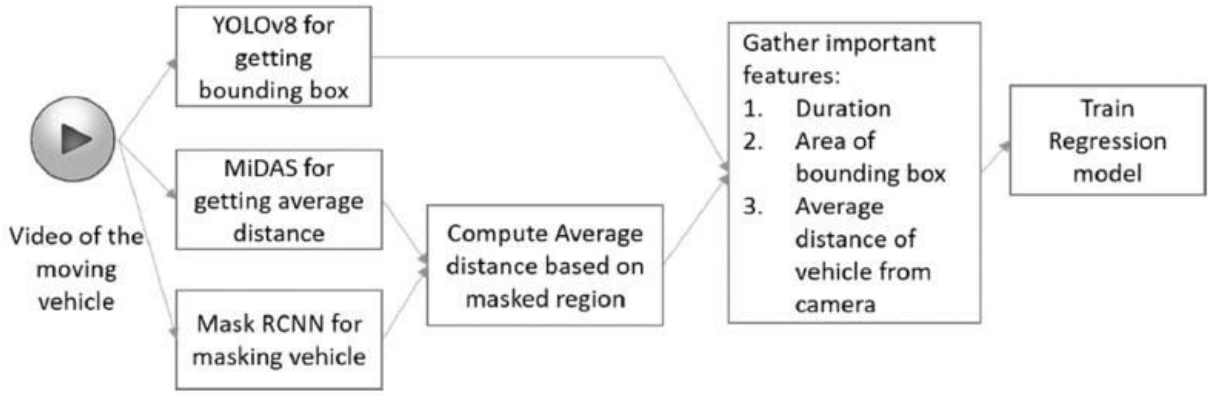

Fig. 1. Flowchart of the model training process

Fig. 2. Object detection with bounding box and associated area (front view)

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

Fig. 3. Object detection with bounding box and associated area (side view)

This was rendered impossible by the bounding box method, as it does not adhere to the object's outline. In order to more precisely determine the average distance of the item of interest alone, object masking was chosen.

Fig. 4. Depth map from the output of MiDAS for the front facing car

This is why, in order to provide a more precise distance estimate, the car was "masked" using a pre-trained RCNN model. Figures 6 and 7 display the results of the vehicle's distance estimate using Mask RCNN and MiDAS, respectively. In these depictions, the masks do their best to mimic the shape of the vehicle. We used the first and last frames of the movie as our starting and ending points to extract the bounding box regions and average vehicle distances for each frame. Then, we used these data points to develop the regression model. Every video had a known and steady vehicle speed, and their durations ranged from 2 to 6 seconds.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

Fig. 5. Depth map from the output of MiDAS for the side facing car.

Fig. 6. Average distance estimation using both MiDAS and Mask RCNN (front view)

Fig. 7. Average distance estimation using both MiDAS and Mask RCNN (side view).

This process began with the construction of a basic linear regression model using the retrieved data. There are two models here; one doesn't include the distance change data from the MiDAS, while the other does.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Having said that, the model's performance was inadequate. Input from MiDAS resulted in an R2 value of about 0.52, whereas input from without it yielded 0.4. This provided strong evidence that MiDAS input played a crucial role in speed estimate. Additionally, the results indicated that there was no linear link between the extracted attributes and the actual speed. Additionally, machine learning based models like SVM or Artificial Neural Network were not applicable because there weren't enough data points. As a result, we attempted non-linear regression using polynomial variables. To generate new features from the preexisting ones, the scikit-learn package in Python was really used for polynomial feature extraction. A dataset was divided into two parts, one for training and one for testing, because overfitting is also possible with polynomial regression. The R2 value increased significantly, and the Adj R2 value was also rather high, while using polynomial regression. Incorporating MiDAS caused the R2 value to soar to 0.81. Object detection studies were conducted using YOLOv5 and SSD in a similar fashion. Table 1 displays the results of the experiments' comparisons.

TABLE I. COMPARATIVE ANALYSIS OF PERFORMANCES OF MODELS

Model (Object Detection)	Regression	Adj R^2	R^2	RMSE
YOLOv8	Linear	0.50	0.52	4.90
	Non-Linear	**0.74**	**0.81**	**2.24**
YOLOv5	Linear	0.45	0.47	5.18
	Non-Linear	0.66	0.76	2.50
SSD	Linear	0.46	0.48	5.28
	Non-Linear	0.60	0.72	2.71

The results with the lowest root-mean-squared errors (RMSEs) were obtained by the YOLOv8 models for object detection, as is evident from the table above. The most important features of the top model are displayed in Figure 8. According to the figure, the most crucial aspect is the change in the average distance of the vehicle from the camera from the first to the last picture. This characteristic is represented by "Diff^1". The square of the difference between the bounding boxes in the first and last frames of the movie follows. It is represented by the symbol "Area Diff^2".

The picture of feature relevance also makes note of other features.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

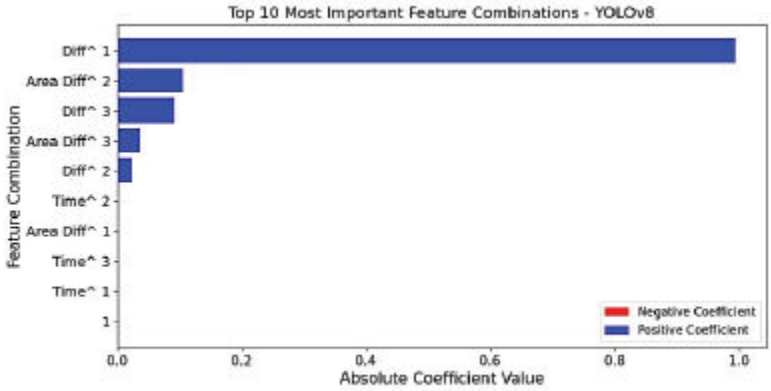

Fig. 8. Feature importance plot of the best model

The correlation between the two sets of numbers is seen in Figures 9 and 10. The output of the ordinary least squares model is seen in Figure 9. A larger variance of error was observed due to the large discrepancy between the model's predictions and the actual values. Figure 10 makes it quite evident, however, that the projected and actual values are closely tracking each other. In addition to the model's performance on the test dataset, the non-linear model's plot displays it with red dots. Also in this scenario, the prediction errors are far lower. Figure 11 displays the sample outputs, which are the final forecasts. The frame number is displayed at the bottom right of each image, and the actual vs. anticipated speed is displayed at the top. There is a definite pattern in the picture about the relationship between the distance from the camera and the accuracy of the vehicle's speed calculation. The precision of the estimated speed declines with increasing distance from the vehicle. The opposite is true: the model's accuracy increases dramatically as the car approaches. It has

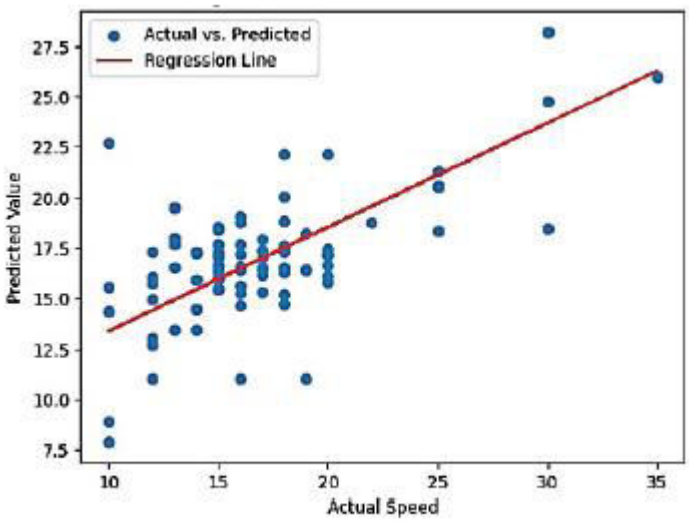

Fig. 9. Actual vs Predicted speed for OLS model.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

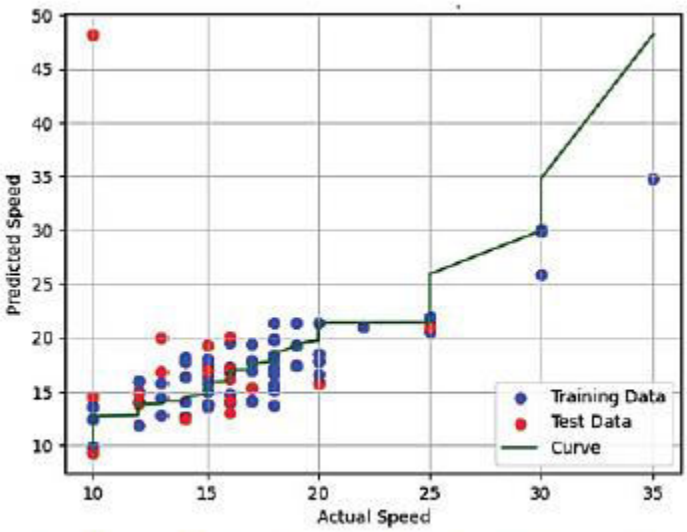

Fig. 10. Actual vs Predicted speed for the non-linear model.

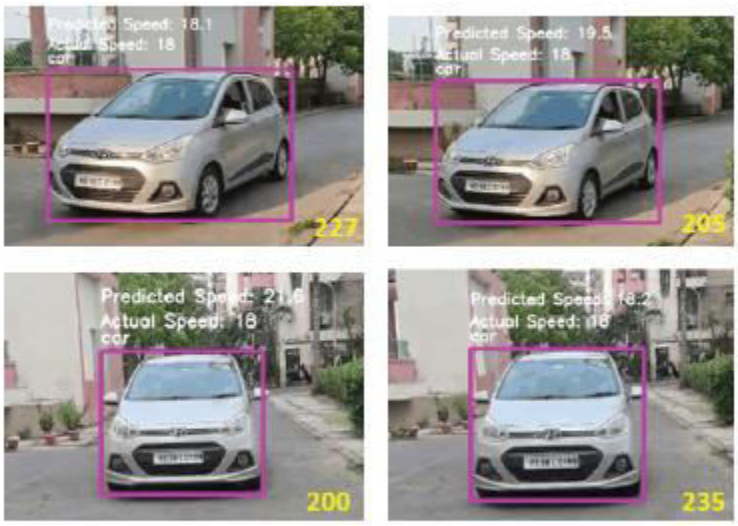

Fig. 11. Sample model output of a moving car in two videos at different frames.

In order to enhance speed estimation models for real-world applications, it is essential to comprehend the link between distance and accuracy. For example, the accuracy of the speed estimate changes significantly based on the car's distance from the camera while looking at the front view in the above image. The anticipated speed was an erroneously higher 21.6 km/h at frame 200, when the automobile was pretty far away, whereas the real speed was 18 km/h. The model fails to provide reliable speed estimates at greater distances from the vehicle, as seen by this mismatch. Nevertheless, by frame 235, the anticipated speed had improved significantly and was nearly identical to the real speed of 18.2 km/h, as the vehicle drew nearer to the camera. It is clear that the model relies on proximity for accurate speed prediction because the accuracy improves significantly as the automobile approaches the camera.

Predictions made from the side are just as indicative of this tendency as those made from the front. At frame 205 of the side view example, the real speed was 18 km/h, whereas the anticipated speed was 19.5 km/h. This first estimate was more than the real speed, just like in the front view case. The anticipated speed, however, was fine-tuned to 18.1 km/h by frame 227, when the camera was getting closer to the vehicle, suggesting better accuracy. A generalizable feature of the approach is shown by this consistency across multiple viewpoints: tighter distances result in improved accuracy in speed prediction. When estimating a vehicle's ultimate speed, it's best to use frames where the vehicle is in the closest proximity to the camera for practical reasons. This method takes use of the fact that a smaller average distance between the vehicle and the camera leads to better accuracy. More accurate and trustworthy results can be produced by the speed estimate model by giving these frames higher priority. This approach guarantees that the model operates at its best, giving precise measurements of speed.

V. CONCLUSION

The current study aimed to estimate the speed of vehicles by making use of deep learning models that have already been trained. There was a noticeable improvement in prediction accuracy with well-executed depth estimation. In addition, the non-linear regression model's predictive ability is significantly improved with more precise bounding box predictions, which in turn improves speed estimates. The YOLOv8 model is able to generate more accurate bounding boxes surrounding the cars. Even though YOLOv8 is fast at conducting analysis, Mask RCNN would be the bottleneck owing to its slower performance compared to YOLOv8, thus it may not be the best solution if speed needs to be calculated at every frame of the video. Therefore, if speed analysis is done every second, this technique might be utilized. Another potential snag occurs when the vehicle is traveling at high speeds, say more than 100 kmph. If that's the case, getting clear shots of the car at fast speeds depends critically on the shutter speed of the camera. Having just one car in the frame was a limitation of the project. Since there would really be more vehicles on the road, the first step in implementing the strategy suggested here would be to implement vehicle monitoring. The study did not include collecting the vehicle's registration number, but that might be changed in the future so that it can be used in conjunction with deep learning to identify license plates for improved police enforcement.

REFERENCES

[1] A. Marode, A. Ambadkar, A. Kale, and T. Mangrudkar, “CARDETECTION USING YOLO ALGORITHM,” International ResearchJournal of Modernization in Engineering Technology and Science, vol.03, pp. 2582–5208, 2021.
[2] X. Zhang, W. Yang, X. Tang, and J. Liu, “A Fast-Learning Method forAccurate and Robust Lane Detection Using Two-Stage FeatureExtraction with YOLO v3,” Sensors, vol. 18, 2018, doi:10.3390/s18124308.

[3] H. Rodríguez-Rangel, L. A. Morales-Rosales, R. Imperial-Rojo, M. A.Roman-Garay, G. E. Peralta-Peñuñuri, and M. Lobato-Báez, "Analysisof Statistical and Artificial Intelligence Algorithms for Real-TimeSpeed Estimation Based on Vehicle Detection with YOLO," Applied Sciences, vol. 12, 2022.

[4] Y. Ming, X. Meng, C. Fan, and H. Yu, "Deep learning for monoculardepth estimation: A review," Neurocomputing, vol. 438, pp. 14–33,2021.

[5] Rao, S. Govinda, R. RamBabu, BS Anil Kumar, V. Srinivas, and P. Varaprasada Rao. "Detection of traffic congestion from surveillance videos using machine learning techniques." In *2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 572-579. IEEE, 2022

[6] J. Zhang, W. Xiao, B. Coifman, and J. P. Mills, "Vehicle tracking andspeed estimation from roadside lidar," IEEE J Sel Top Appl Earth ObsRemote Sens, vol. 13, pp. 5597–5608, 2020.

[7] P. Misans and M. Terauds, "CW doppler radar based land vehicle speedmeasurement algorithm using zero crossing and least squares method,"in 2012 13th Biennial Baltic electronics conference, 2012, pp. 161–164.

[8] Agrawal, K. K. ., P. . Sharma, G. . Kaur, S. . Keswani, R. . Rambabu, S. K. . Behra, K. . Tolani, and N. S. . Bhati. "Deep Learning-Enabled Image Segmentation for Precise Retinopathy Diagnosis". *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 12s, Jan. 2024, pp. 567-74, https://ijisae.org/index.php/IJISAE/article/view/4541.

[9] S. Hua, M. Kapoor, and D. C. Anastasiu, "Vehicle tracking and speedestimation from traffic videos," in Proceedings of the IEEE Conferenceon Computer Vision and Pattern Recognition Workshops, 2018, pp.153–160.

[10] D. Bell, W. Xiao, and P. James, "Accurate vehicle speed estimationfrom monocular camera footage," ISPRS Annals of thePhotogrammetry, Remote Sensing and Spatial Information Sciences,vol. 2, pp. 419–426, 2020.

[11] Samota, H. ., Sharma, S. ., Khan, H. ., Malathy, M. ., Singh, G. ., Surjeet, S. and Rambabu, R. . (2024) "A Novel Approach to Predicting Personality Behaviour from Social Media Data Using Deep Learning", *International Journal of Intelligent Systems and Applications in Engineering*, 12(15s), pp. 539–547. Available at: https://ijisae.org/index.php/IJISAE/article/view/4788

[12] R. Birkl, D. Wofk, and M. Müller, "Midas v3. 1--a model zoo for robustmonocular relative depth estimation," arXiv preprintarXiv:2307.14460, 2023.