



Article

Deep Learning and YOLOv3 Systems for Automatic Traffic Data Measurement by Moving Car Observer Technique

Marco Guerrieri ^{1,*} and Giuseppe Parla ²

¹ Department of Civil, Environmental and Mechanical Engineering, University of Trento, Via Mesiano 77, 38123 Trento, Italy

² ISMETT, Via E. Tricomi 5, 90127 Palermo, Italy; gparla@ismett.edu

* Correspondence: marco.guerrieri@unitn.it

Abstract: Macroscopic traffic flow variables estimation is of fundamental interest in the planning, designing and controlling of highway facilities. This article presents a novel automatic traffic data acquisition method, called MOM-DL, based on the moving observer method (MOM), deep learning and YOLOv3 algorithm. The proposed method is able to automatically detect vehicles in a traffic stream and estimate the traffic variables flow q , space mean speed vs and vehicle density k for highways in stationary and homogeneous traffic conditions. The first application of the MOM-DL technique concerns a segment of an Italian highway. In the experiments, a survey vehicle equipped with a camera has been used. Using deep learning and YOLOv3 the vehicles detection and the counting processes have been carried out for the analyzed highway segment. The traffic flow variables have been calculated by the Wardrop relationships. The first results demonstrate that the MOM and MOM-DL methods are in good agreement with each other despite some errors arising with MOM-DL during the vehicle detection step due to a variety of reasons. However, the values of macroscopic traffic variables estimated by means of the Drakes' traffic flow model together with the proposed method (MOM-DL) are very close to those obtained by the traditional one (MOM), being the maximum percentage variation less than 3%.

Keywords: deep learning; traffic flow; moving car observer technique; traffic data acquisition



Citation: Guerrieri, M.; Parla, G. Deep Learning and YOLOv3 Systems for Automatic Traffic Data Measurement by Moving Car Observer Technique. *Infrastructures* **2021**, *6*, 134. <https://doi.org/10.3390/infrastructures6090134>

Academic Editors: Reza Haghani, Rami Hawileh and Paolo S. Valvo

Received: 23 August 2021

Accepted: 11 September 2021

Published: 18 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Macroscopic traffic flow variables estimation is very important in the planning, designing and controlling processes of highway and road networks. In addition, Intelligent Transportation Systems (ITS) require that the observation and the control of traffic demand and flow rate be performed in real-time with more accuracy and reliability than in the past [1].

Traffic data can be collected via several methods and technologies, including the “fixed spot measurement” (i.e., inductive-loops, pneumatic tubes, detectors video cameras, etc.) and the “probe vehicle data” systems (i.e., floating car data FCD) both widely used (Table 1). In fixed spot measurement methods, for the vehicle detection and counting processes can be applied several algorithms based on Deep Learning [2–5].

An Alternative method employed to detect the flow rate values on a certain highway section is the “moving car observer method” (MOM), also called “moving car observer” (MCO) established by Wardrop and Charlesworth [6]. In this technique, several runs of a test vehicle (observer or survey vehicle) are carried out traveling the test vehicle with and against the one-way traffic stream of interest.

In the traditional MOM, the fundamental traffic information (i.e., the quantity of light and heavy vehicles and the journey times) are generally recorded by three human observers at least that are inside the survey vehicle.

Table 1. Methods and technologies for traffic flow variables measure.

Method	Technologies	Measured Traffic Variables
fixed spot measurement	<ul style="list-style-type: none"> • hand tally • pneumatic tubes • inductive loop • microwave, radar • photocell • ultrasonic • video camera 	<ul style="list-style-type: none"> • volume • flow rate • time headway • instantaneous speed (density is calculated)
Measurement over a short road section (less than about 10 m)	<ul style="list-style-type: none"> • pneumatic tubes • paired detectors: inductive loops, microwave beams • video camera 	<ul style="list-style-type: none"> • volume • speed • time headway • occupancy (density is calculated)
Measurement over a road section (at least 0.5 km)	<ul style="list-style-type: none"> • aerial photography • cameras installed on tall buildings or poles 	<ul style="list-style-type: none"> • density • speed • travel time
Moving car observer method (MOM)	<ul style="list-style-type: none"> • floating vehicle with human observers (hand tally) 	<ul style="list-style-type: none"> • travel time • speed • flow rate • density
Moving car observer method with Deep learning (MOM-DL)	<ul style="list-style-type: none"> • floating vehicle equipped with a video camera 	<ul style="list-style-type: none"> • travel time • speed • flow rate • density

This article proposes a novel real-time automated MOM in which the vehicle classification and the counting processes are obtained by Computer Vision and deep learning approaches.

Contrarily from the traditional MOM, the proposed one, denoted with the term MOM-DL (Moving car observer with Deep Learning method) does not necessitate human observers, thus reducing measurement expenditures while growing speed of the vehicles counting, and the safety of road users.

Table 1 summarizes the main methods, including the proposed MOM-DL technique, and the required technologies for measuring the traffic flow variables.

MOM-DL method is able to estimate the macroscopic traffic flow variables in each section of a single road infrastructure or in several sections of a road network with remarkable benefits with respect to the traditional Moving car observer method. The proposed automatic traffic data acquirement method has been verified, calibrated and validated by experiments conducted in a two-lane undivided highway.

Counting processes are investigated via a survey vehicle equipped with a calibrated camera. Traffic video recordings, with a resolution of 1280 × 720 pixels, have been analyzed using a workstation with Intel(R) Core(TM) i7-4510 CPU @ 2.00 Hz 2.60 GHz—Memory RAM 32 GB, Windows 10 Home. Experiment outcomes, achieved by means of the suggested technique (MOM-DL), are in good accordance with the data deduced by the conventional moving observer method (MOM).

The outline of the paper is as follows. Section 2 explains the related work of object detection and recognition systems based on the deep learning approach and YOLO v3 algorithm. Section 3 briefly explains the “moving car observer method” (MOM) and the proposed technique MOM-DL. The experiments are presented in Section 4, and results and discussions are presented in Section 5. Finally, conclusions are proposed in Section 6.

2. Deep Learning and YOLOv3 Algorithm

2.1. General Considerations

In recent times, artificial intelligence and deep learning (DL) have shown several potential applications in many real-life areas including target detection [7]. In deep learning, a computer model learns to perform classification tasks using different types of information such as texts, sounds or images. The models implemented in deep learning are formed using a large number of tagged data and neural network architectures that contain various layers [8]. The word “deep” refers to the number of hidden layers in the neural network.

In computer vision (CV) applications, the most important DL architectures are artificial neural networks (ANNs), convolutional networks (CNNs) and generative adversarial networks (GANs) [9]. In computer vision, image classification problems are the most basic applications for CNNs. Object detection systems like YOLO (You Only Look Once), SSD (single-shot detector) and Faster R-CNN, not only classify images but also can locate and detect each object in images that contain multiple objects [9].

In 2016, Redmon and his team proposed the YOLO convolutional neural network model that can complete end-to-end training [10]. In the same year, an advanced version of YOLO, YOLOv2, was developed, which can maintain high detection accuracy at a high speed and has a high advantage in real-time image processing [11,12].

With the popularization of artificial intelligence technology, unmanned vehicles and automatic robots are more and more widely used. In the future, emerging technologies based on artificial intelligence technology such as automated vehicles AVs and connected automated vehicles CAVs will be part of transportation systems.

AVs and CAVs need to perform operations based on the infrastructure and environmental scenario they are facing, including stopping at intersections and pass walk areas, performing acceleration, deceleration and turning maneuvers, etc. Consequently, the development and application of traffic scene, vehicle and vulnerable user classification has become a very significant research topic in intelligent driving systems, which has a far-reaching influence on transportation and highway engineering.

In recent years, object detection algorithms have made great advances. In this field, the most widely used algorithms can be classified as follows [13]:

- (a) R-CNN system algorithm in which the Heuristic method or convolutional neural network are applied to produce alternative regions first, and then classification and regression are used on the alternative regions;
- (b) end-to-end algorithms (i.e., You Look Only Once “Yolo” and Single-Shot Detector “SSD”), that only use a convolutional neural network to directly predict the category and location of different objects [12].

YOLO and SSDs are one-stage detectors and conduct target classification and localization at the same time. One of the most important innovations of YOLOv2, based on a 31-layer neural network, is the concept of the anchor box obtained by k-means, instead of the traditional bounding box. Anchor box increases the probability of an object being identified from 81% to 88% [13].

YOLOv3 is the latest version of YOLO and does not require a region proposal network (RPN) and directly performs regression to detect targets in the image [14]. In brief, YOLOv3 includes 53 convolutional layers and 23 residual layers as shown in Figure 1. YOLOv3 has shown significant advancement in real-time object detection, especially in the detection of smaller objects. Consequently, YOLOv3 is used for our vehicles detection system.

As explained in [14], 1×1 , $3 \times 3/2$ and 3×3 convolution kernels of three sizes are applied in the convolutional layers to sequentially extract image features, ensuring the model has remarkable classification and detection performances. The remaining layers guarantee the convergence of the detection model [14].

In order to detect several areas of the same object at the same time, YOLOv3 fuses three feature maps of different scales (52×52 , 26×26 and 13×13) by three time down sampling.

In this research, YOLOv3 is adopted because the YOLO family (i.e., YOLOv1, YOLOv2, YOLOv3) is a series of end-to-end deep learning models planned and designed for fast

real-time object detection [9] and the quick vehicles detection is essential for the proposed MOM-DL technique.

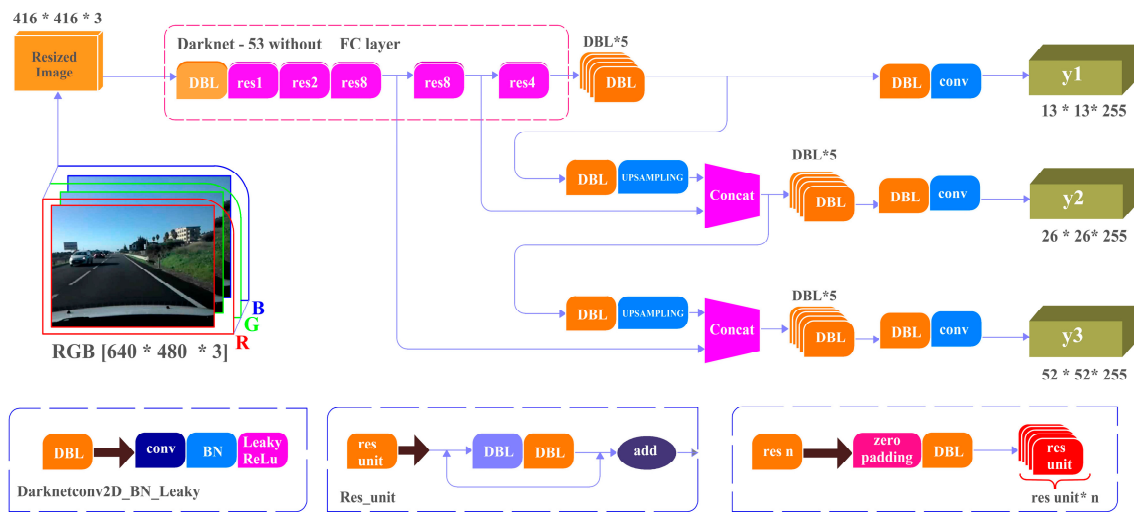


Figure 1. YOLOv3 Network structure (adapted from [15]).

2.2. YOLO V3 Algorithm

The network structure of YOLOv3 model uses the Darknet-53 network model that includes 5 residual modules and 53 convolutional layers (Figure 2).

As well explained in [16] the Darknet-53 network can “performed 5 down-sampling of the image, each of which has a sampling step size of 2 and a maximum step size of 32. Then the image is subjected to 32-fold down-sampling, 16-fold down-sampling, and 8-fold down-sampling processing to obtain 3 target detections with scale differences, Then the image is subjected to 32-fold down-sampling, 16-fold down-sampling, and 8-fold down-sampling processing to obtain 3 target detections with scale differences, the feature maps are respectively 13×13 , 26×26 , 52×52 ; The first feature map is suitable for detecting large targets, the second feature map tends to detect medium targets, and the last one is mostly used for small targets. Finally, the output feature map is sampled and feature fusion is performed, and finally the detection task of the target is completed”.

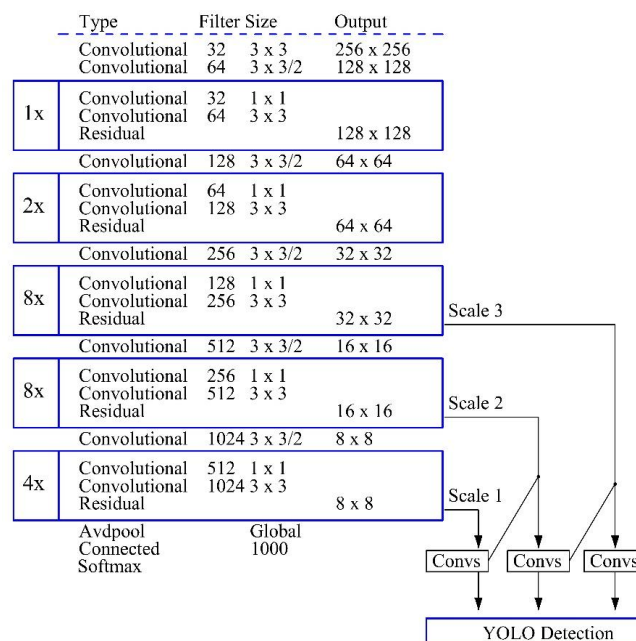


Figure 2. Darknet-53 structure diagram in YOLOv3 (adapted from [17]).

During the training process, the analyzed image is subdivided into $S \times S$ grids by the YOLOv3 model and each grid forecasts if the center of the object falls into its inner or not. In the first case, the grid predicts B detection bounding boxes and the $\text{Conf}(\text{Object})$ (i.e., the confidence of each box). In the second case, the candidate box confidence for the nonexistence of a target is imposed to zero. The previous conceptual and logical phases can be formalized with Equations (1)–(3) [9,18]:

$$\text{Conf}(\text{Object}) = \text{Pr}(\text{Object}) \times \text{IoU}_{\text{pred}}^{\text{truth}} \tag{1}$$

$$\text{Pr}(\text{Object}) = \begin{cases} 0 & \text{no target in the cell} \\ 1 & \text{there are targets in the cell} \end{cases} \tag{2}$$

$$\text{IoU}_{\text{pred}}^{\text{truth}} = \frac{B_{\text{groundtruth}} \cap B_{\text{predicted}}}{B_{\text{groundtruth}} \cup B_{\text{predicted}}} \tag{3}$$

where $\text{Conf}(\text{Object})$ is the confidence of the candidate box corresponding to the cell and $\text{IoU}_{\text{pred}}^{\text{truth}}$ is the intersection over union (ratio of the intersection area of the prediction box to the actual frame and the area of the union).

The $\text{IoU}_{\text{pred}}^{\text{truth}}$ measure evaluates the overlap between two bounding boxes: the ground truth bounding box ($B_{\text{ground truth}}$), that is, the hand-labeled bounding box created during the labeling process and the predicted bounding box ($B_{\text{predicted}}$) from the considered and proposed model. Each detected bounding box comprises the following parameters:

- (x, y) : position of the center of the detection bounding box relative to its parent grid;
- (w, h) height and width of the detection bounding box;
- $\text{Pr}(\text{Class}_i | \text{Object})$ probability of C categories: is the probability that the center of the i -th object falls into the grid.

Lastly, a tensor of the $S \times S (B \times 5 + C)$ dimension is calculated.

2.3. Loss Calculation

YOLO predicts multiple bounding boxes per grid cell. To compute the loss for the true positive, it is required that one of them is responsible for the object. Therefore, we selected the one with the highest IoU (intersection over union) with the ground truth.

For calculation of loss, in YOLO, the sum-squared error between the predictions and the ground truth is used. The loss function comprises the classification loss, the localization loss (errors between the predicted boundary box and the ground truth) and the confidence loss (the objectness of the box) [19] as follows:

- Classification loss: if an object is detected, the classification loss at each cell is the squared error of the class conditional probabilities for each class [19]:

$$\sum_{i=0}^{S^2} I_{ij}^{\text{obj}} (p_i(c) - \hat{p}_i(c))^2 \tag{4}$$

where:

$I_{ij}^{\text{obj}} = 1$ if an object appears in the cell I , otherwise is 0;

$\hat{p}_i(c)$ the conditional class probability for class c in cell i .

- Localization loss: evaluates the errors in the predicted boundary box locations and sizes. It is only counted the box responsible for detecting the object [19].

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} [(x - \hat{x}_i)^2 + (y - \hat{y}_i)^2] + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} [(\sqrt{\omega_i} - \sqrt{\hat{\omega}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] \tag{5}$$

where:

$I_{ij}^{\text{obj}} = 1$ if in the j th boundary box in cell i is responsible for detecting the object, otherwise 0;

λ_{coord} increase the weight for the loss in the boundary box coordinates.

Confidence loss: if an object is detected in the box, the confidence loss (measuring the objectness of the box) is [19]

$$\sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} (C - \hat{C}_i)^2 \tag{6}$$

where:

\hat{C}_i stands for the box confidence score of the box j , in the cell i ;

$I_{ij}^{obj} = 1$ if the j th boundary box in the cell i is responsible for detecting the object, otherwise 0.

If an object is not detected in the box, the confidence loss is [19]

$$\lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{noobj} (C - \hat{C}_i)^2 \tag{7}$$

where

λ_{noobj} is the complement of I_{ij}^{obj} ;

\hat{C}_i is the box confidence score of the box j in the cell i ;

λ_{noobj} weights down the loss when detecting background.

The final loss adds localization, confidence and classification losses together [19], as follows:

$$\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} [(x - \hat{x}_i)^2 + (y - \hat{y}_i)^2] + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} [(\sqrt{\hat{w}_i} - \sqrt{\hat{h}_i})^2 + \sqrt{\hat{h}_i} - \sqrt{\hat{h}_i}] + \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} (C - \hat{C}_i)^2 + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{noobj} [(C - \hat{C}_i)^2 + \sum_{i=0}^{S^2} I_{ij}^{obj} (p_i(c) - \hat{p}_i(c))^2] \tag{8}$$

In YOLO v3, the method of predicting the bounding box (cfr. Figure 3) is given by Equation (9):

$$\begin{cases} b_x = \sigma(t_x) + c_x \\ b_y = \sigma(t_y) + c_y \\ b_w = p_w e^{t_w} \\ b_h = p_h e^{t_h} \end{cases} \tag{9}$$

where t_x, t_y, t_w and t_h are the predicted outputs of the model, which represent the relative position coordinates of the center of the bounding box and the relative width and height of the bounding box. c_x and c_y represent the net, and p_w and p_h are the width and height of the predicted front bounding box. Finally b_x, b_y, b_w and b_h are the true coordinates of the center of the bounding box and the true width and height of the bounding box obtained after prediction.

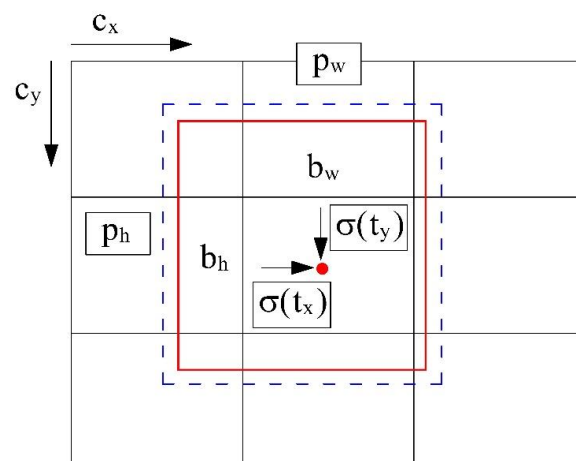


Figure 3. Bounding Box with dimensions priors and location prediction (adapted from [17]).

The performance of a certain object detector can be measured mainly by the following metrics [20]:

- Average mean precision (mAP) that is the integral over the precision $p(o)$:

$$mAP = \int_0^1 p(o) do \tag{10}$$

where $p(o)$ is the precision of the object detection.

- Frame per second (FPS) to measure detection speed (number of images processed every second);
- Precision-recall curve (PR curve) in which Precision and Recall are calculated as follows:

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

- Accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

The symbols TP, FN and FP stand for True Positive, False Negative and False Positive respectively.

- Another parameter frequently used is the F1-score that is the harmonic mean of the precision and recall values:

$$F1 = \frac{2}{\frac{1}{precision} + \frac{1}{recall}} = \frac{2TP}{2TP + FP + FN} \tag{14}$$

F1-score can assume values in the range [0, 1]: if F1-score = 1 it results a perfect classifier. With reference to the semantic segmentation of images, F1-score indicates the proximity between the traced contour and the segmented contour.

3. Automatic Traffic Data Measurement Using Moving Observer Method

The moving observer method (MOM) was developed by Wardrop and Charlesworth [6] for macroscopic traffic variables measurements. Both the space mean speed vs. flow rate q measurements are obtained simultaneously through this method. This method requires a survey vehicle that travels in both directions on the analyzed highway. This procedure needs numerous runs of a survey vehicle done traveling in the same direction and against the traffic stream of interest. For each run (i.e., experiment) the human observers in the survey vehicle measure several traffic parameters, including the number of opposing vehicles met, the number of vehicles that the test vehicle overtook, the number of vehicles overtaking the test vehicle, the test vehicle mean speed, the length and the travel time of the run [6,21].

Equations (15)–(17) are used to calculate both the space mean speed and traffic flow for one direction of travel. In addition, vehicle density k can be estimated with Equation (18).

$$q = \frac{x + y}{(t_a + t_w)} \tag{15}$$

$$t_s = t_w - \frac{y}{q} \tag{16}$$

$$v_s = \frac{L}{t_s} \tag{17}$$

$$k = \frac{q}{v_s} \quad (18)$$

where:

- q is the estimated traffic flow on the lane in the direction of interest;
- v_s is the space mean speed in the direction of interest;
- k is the vehicle density in the direction of interest;
- x denotes the number of vehicles traveling in the direction of interest that are met by the survey vehicle while traveling in the opposite direction;
- y is the net number of vehicles that overtake the survey vehicle while traveling in the direction of interest (overtake test car—overtaken by test car);
- t_a is the travel time taken for the trip against the stream;
- t_s indicates the estimate of mean travel time in the direction of interest;
- t_w is the travel time of the test vehicle with the stream;
- L is the length of the highway segment under analysis and corresponds to the distance traveled by the survey vehicle.

According to [6], traffic flow values estimated by MOM may result in a small degree of bias. Counting processes are not affected by small deviations in the speeds of the test of the observed vehicles. Furthermore, no significant differences were found between macroscopic traffic flow variables values estimated by MOM and fixed spot measurement. Nevertheless, systematic errors may occur if the test vehicle speed v_w and the space mean speed v_s of the traffic streams have very similar values [6]. In the case of one-way roads, the observer vehicle is required to travel along the segment of interest at least with two speed values [22–24]. In accordance with [25], in order to guarantee reliable results with the use of MOM the choice of survey highway segment should stick with the following basic conditions:

- Homogeneity: homogeneous in geometric characteristics (horizontal and vertical alignment, lanes and shoulders width, etc.) throughout the whole segment of length L ;
- Intersections: there is no at-grade intersection within the highway segment or within at least 250 m of its endpoints;
- Speed limits: no sub-segment contained within it, or within 250 m of its endpoints, any speed restriction respect to the legal speed limit;
- Roadworks: there are no roadworks;
- Length: the segment length should be in the range of 1–5 km.

The suggested method MOM-DL, based on Computer Vision and Deep learning, is able to automatically detect, track and count along the analyzed highway segment the vehicles identified with “ x ” and “ y ” previously defined and classify them into different categories (motorbikes, light vehicles, heavy vehicles, etc.). The process is subdivided into subsequent steps: Vehicle Detection, Vehicle Classification and Recognition, Labeling, Tracking, Counting process.

In the counting process, first, the traffic variables x and y , t_a and t_w are evaluated and, lastly, the flow (q), the space mean speed (v_s) and density (k) are obtained using Equations (15)–(18). It is worth underlining that in a given video frame sequence the bounding box perimeter and surface of each detected vehicle change continuously over the time instants. The growth over time of the bounding box perimeter and area represents a reduction in the space headways between the observed vehicle (i.e., leader vehicle) and the survey vehicle (i.e., follower vehicle). Obviously, after an overtaking maneuver, the bounding box of the observed vehicle disappears in the video frame sequence. Then it is possible to count the number “ y ” of vehicles that overtake the survey vehicle when it travels the highway in the same direction of the traffic stream that should be examined. Correspondingly, the number “ x ” of vehicles met by the test vehicle traveling the highway segment in the opposite direction to the traffic stream of interest may be automatically measured.

These counting processes were implemented in Matlab R2020b using YOLOv3 and were applied in a case study (cfr. Section 4). The first results demonstrate that the MOM-DL is very accurate though various false positives (FP) and false negatives (FN) can be found in

the detection step. This is because the test vehicle, and in turn the video camera, is subject to vibrations [26] caused by road pavement surface irregularities (including holes, bumps, and uneven pavement edge), lighting reflections and adverse environmental conditions (rain, fog darkness, etc.).

4. Experiments

The proposed method was applied to a segment of the Italian two-lane single carriageway SS640 highway. The analyzed highway segment of length L ($L = 1.1$ km), is characterized by uninterrupted flow conditions. Furthermore, this segment satisfies the basic conditions explained in the previous Section 3 concerning the homogeneity, intersections, speed limits, roadworks and length. The endpoints of the segment are denoted by sections A and B (Figure 7).

In accordance with the moving observer method, the survey vehicle was driven several times in the same direction and in the opposite direction with respect to the traffic stream that was to be measured (flow q). The survey vehicle speed was prefixed in the range 60 ± 5 km/h depending on the traffic condition. In the year 2021, 35 round trips for each direction were performed in the analyzed highway segment.

The videos of the traffic streams were recorded by a survey vehicle equipped with a calibrated camera (Figure 4). The first research phase was the camera calibration. Camera calibration aims at establishing two sets of parameters: intrinsic and extrinsic. The intrinsic parameters correct lens distortions while the extrinsic ones determine the spatial offset of the camera. The Zhang algorithm [27] determines the extrinsic parameters of the system. For the calibration processes, 64 different images were considered using a set of different photos as shown, for example, in Figure 5. Figure 6 shows the extrinsic parameter visualization obtained by the calibration processes. The calibrated model was then validated by several tests concerning the comparison between estimated and measured distances of the prefixed objects with respect to the camera lens. A similar technical approach can be applied in several fields of highway and transportation engineering [28,29].



Figure 4. Survey vehicle and traffic recording system.

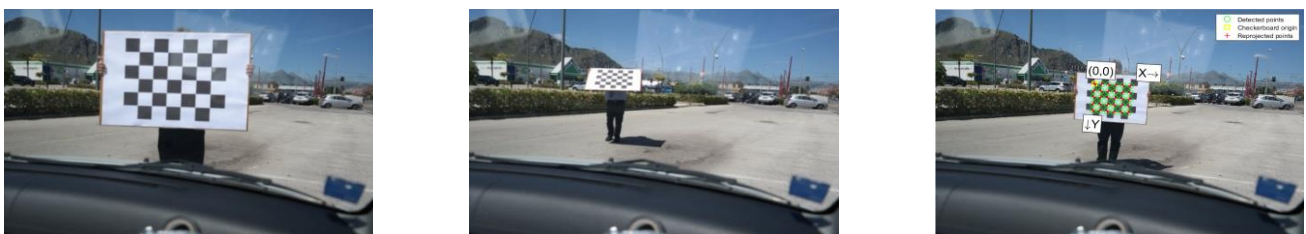


Figure 5. Camera calibration process.

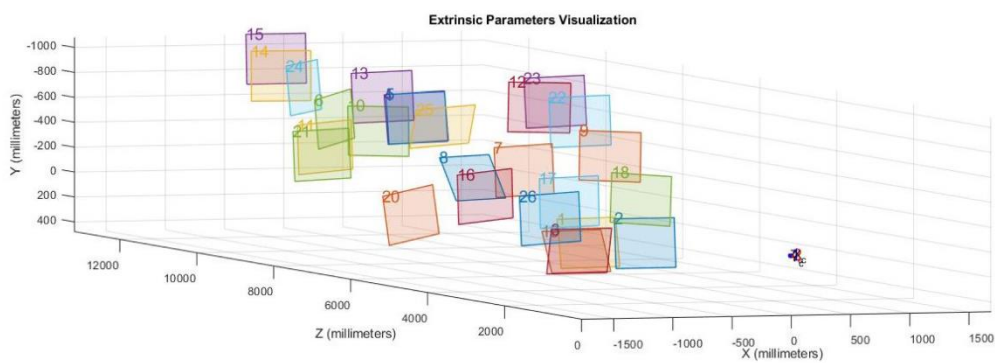


Figure 6. Extrinsic parameters visualization.

By using Equations (15)–(18), for each run, the following parameters were estimated with the method MOM-DL:

- The number y of vehicles that overtake the survey vehicle when it travels in the direction from A to B (see Figure 7 and Figure 12);
- The number x of vehicles traveling in the direction of interest that are met by the survey vehicle while traveling in the opposite direction, from B to A (see Figure 7 and Figure 13);
- The travel times t_a and t_w spent for crossing the segment in the two directions (from A to B and from B to A).

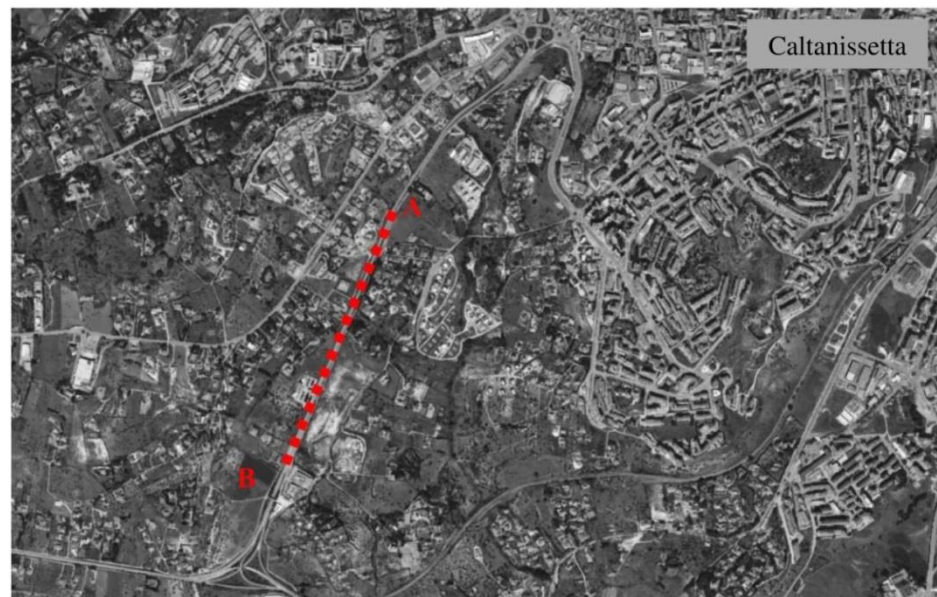


Figure 7. Analyzed segment of the highway SS 640.

Training of the Neural Networks: Main Results

As explained in the previous Section 2.2 the Darknet-53 is a framework to train neural networks. Darknet-53 is open-source and is written in C and CUDA and serves as the basis for YOLOv3 [18].

Darknet-53 structure is summarized in Figure 2 [10,17]. In general, in Darknet-53, the weights of the custom detector are saved for every 100 until 1000 iterations, and it continues to be saved for every 10,000 iterations until it reaches the maximum batches [17]. A pre-existing vehicle dataset, consisting of 652 images published by [30–32] was used in this study. The pre-trained network can classify images into numerous object categories, such as cars, heavy vehicles, motorbikes, pedestrians, animals, etc. Also, an additional

set of front images of private and public vehicles was added to the pre-existing vehicles dataset. The pre-existing vehicles dataset (including light and heavy vehicles, motorbikes, pedestrians and other users) in Darknet-53 was split into 70% for training and 30% for testing the neural networks.

As a result, the network learned feature-rich representations for a wide range of vehicle images. During the training process phase, data augmentation techniques (cropping, padding, flipping, etc.) were used in order to prepare the large neural networks. Then, a bounding box labeling tool [33] was applied to manually detect and recognize vehicles for the object to be detected [17]. Regarding the processing time of the training phase, consider, for example, the following parameters: 8 Epoch (1 Epoch = 25 iterations), learning rate = 0.001, L2 regularization factor = 0.0005, penalty threshold = 0.5, the result is a Time Elapsed = 00:40:20 h. The outcomes of this phase and the class label consist of four points of the position coordinate. The label was converted into YOLO format and the tool changed the values to a format that the training algorithm YOLOv3 can employ.

Figures 8 and 9 show the training process consisting of 200 iterations. The Accuracy (Equation (13) and Figure 8), the Loss (Equation (8) and Figure 9) and the Precision (Equation (12) and Figure 10) values demonstrate that the proposed training model is able to detect the vehicles with high accuracy.

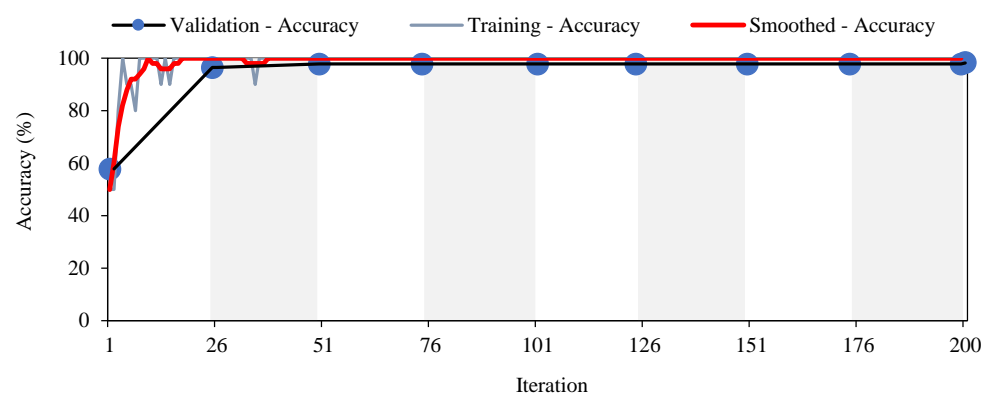


Figure 8. Accuracy evolution related to the number of iterations.

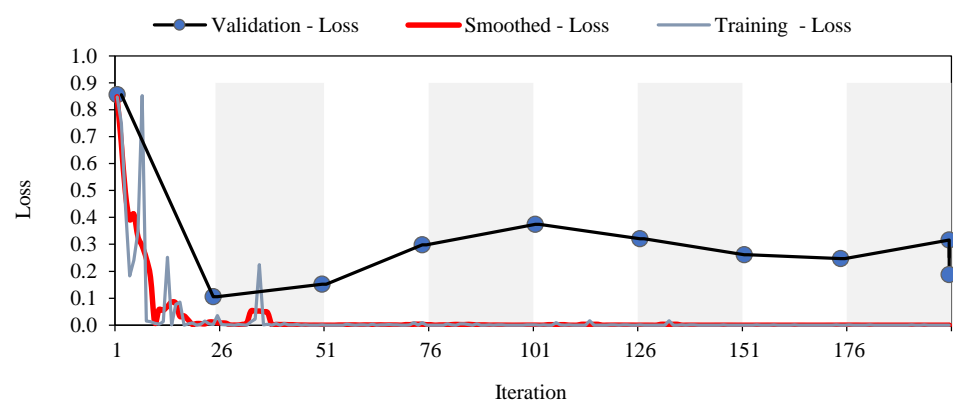


Figure 9. Loss evolution related to the number of iterations.

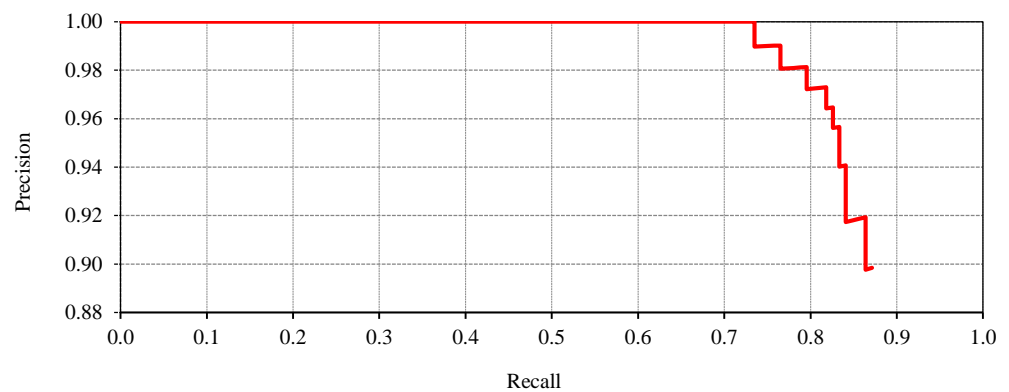


Figure 10. YOLOv3 precision-recall curve.

5. Results and Discussions

In computer vision, the tracking phase allows the detection of the same object (i.e., vehicle) in all its locations that may be identified in the consecutive frames of the recorded video. The phases employing the tracking algorithm are exemplified in Figure 11. During each test, in the sequence of video frames both the perimeter and surface of the bounding box relating to each vehicle changes over time. Therefore, the increase over time of this surface denotes the space headway decrease between the observed vehicle and the survey vehicle. Moreover, after an overtaking maneuver, the bounding box of the overtaken vehicle disappears [20,34]. Consequently, the traffic variable “y” (i.e., vehicles which overtaking the survey vehicle when it traveling the segment A-B, Figure 12) can be measured. Correspondingly, the number “x” of vehicles met by the survey vehicle while it was traveling in the opposite direction (i.e., against the traffic stream of flow q) was automatically measured, Figure 13. Then, for each of the thirty-five runs performed in the experiments, the macroscopic traffic variables (q ; v_s ; k) were automatically obtained by Equations (15)–(18).

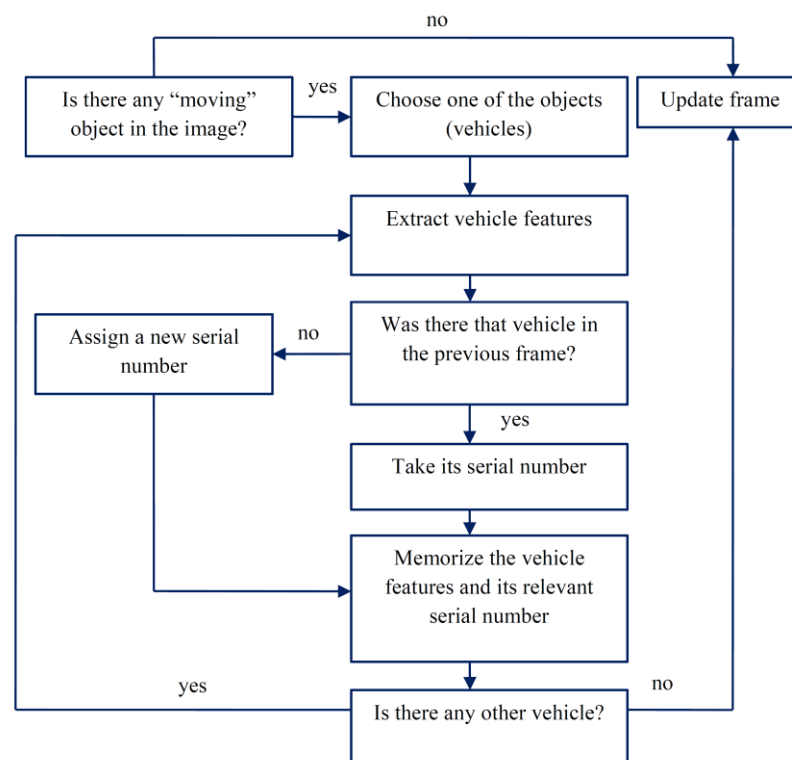


Figure 11. Flow chart of the tracking algorithm.

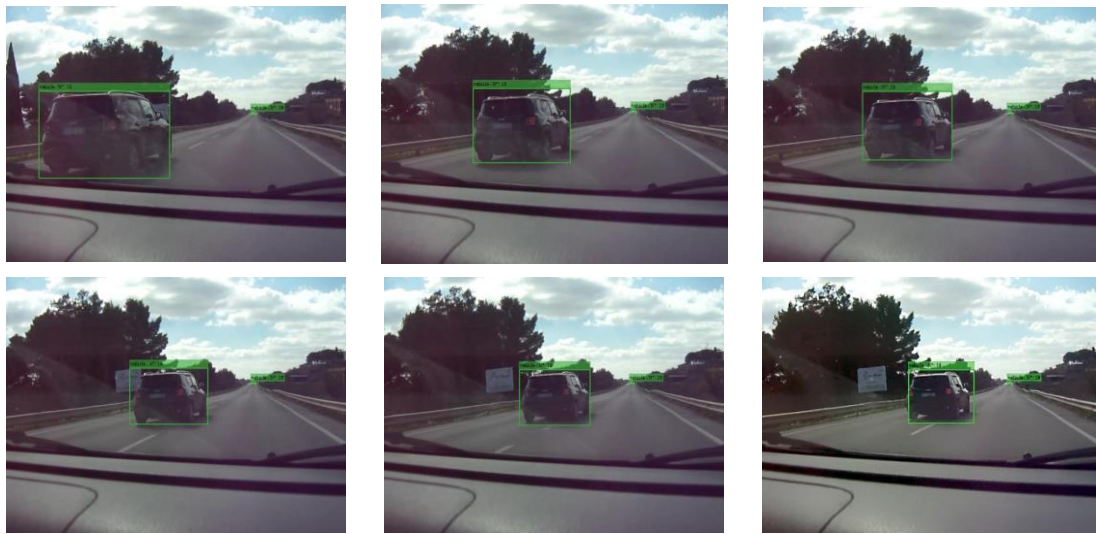


Figure 12. Examples of vehicles detection and tracking phases for the estimation of the traffic variable “y”. (direction from A to B, cfr. Figure 7).

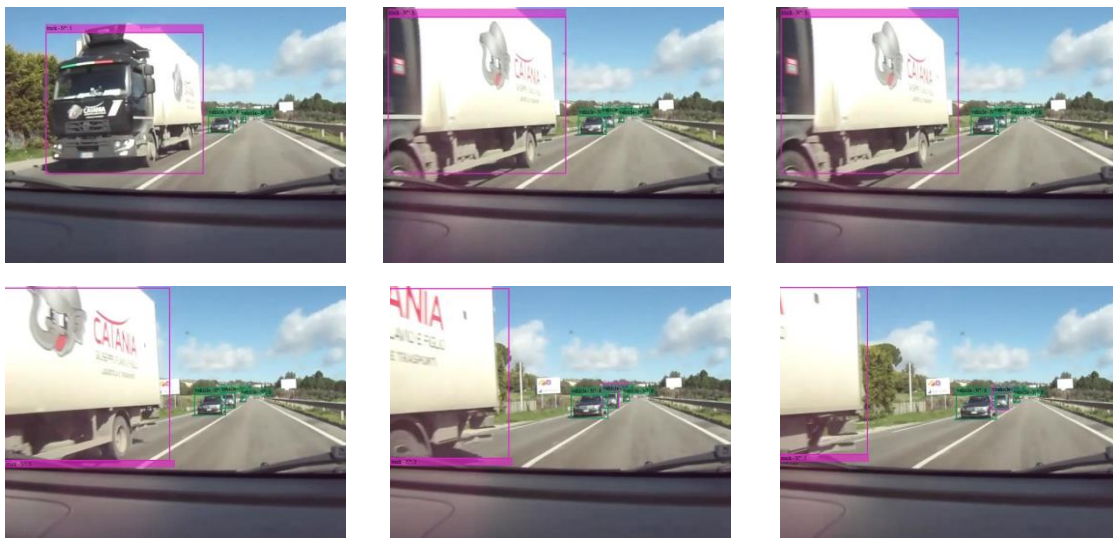


Figure 13. Examples of vehicles detection and tracking phases for the estimation of the traffic variable “x”. (direction from A to B, cfr. Figure 7).

Different vehicle types were detected [35] and homogenized to passenger car unit (pcu) by means of the HCM 2016 coefficients (flows were then expressed in pcu/h).

In order to estimate capacity, free-flow speed and other fundamental traffic variables, first the main traffic flow models $v = v(k)$ were considered. Among these models, the bell-shaped curve model proposed by Drake, Schofer and May [36–38] (in short, the “May model” or “Drake model”) was adopted because it proved to be the best in interpreting the available experimental data. May’s equation, after logarithmic transformations, is as follows:

$$\ln(v) = \ln(v_f) - \frac{k^2}{2 \cdot k_c^2} \tag{19}$$

or

$$V_1 = a + bD_1 \tag{20}$$

in which $V_1 = \ln(v)$; $a = \ln(v_f)$; $b = 1/(2 k_c^2)$; $D_1 = k^2$. Where v_f denotes the free-flow speed and k_c the critical density, that is, the vehicle density associated with the capacity. By means of the fundamental flow relation $q = k \cdot v$, it results in the following equations:

$$q = v \cdot \sqrt{\frac{\ln \frac{v_f}{v}}{\frac{0.5}{k_c^2}}} \tag{21}$$

$$q = v_f \cdot k \cdot e^{-\frac{1}{2}(\frac{k}{k_c})^2} \tag{22}$$

which allow the two additional flow relations to be traced: $v_s = v_s(q)$ and $q = q(k)$. Finally, the flow models were calibrated. By observing Equations (21) and (22), it can be realized that the two parameters needed to be assessed are the free-flow speed v_f and the critical density k_c .

On the basis of scatter plots ($k; \ln(v_s)$) the model calibration parameters v_f and k_c were deduced as shown in Figure 14 for both traffic data acquirement methods (i.e., MOM and MOM-DL). For instance, in the case of MOM-DL method, the Figure 14b shows the speed –density scatter points ($k; \ln(v_s)$). In this case, it results: $v_f = \exp(a) = \exp(4.2745) = 71.84$ km/h; $k_c = \sqrt{\frac{1}{-2 \cdot b}} = \sqrt{\frac{1}{-2 \cdot (-0.0006)}} = 28.87$ pcu/km/lane. Therefore, the calibrated relation $q = q(k)$ is: $q = 71.84 \cdot k \cdot e^{-\frac{1}{2}(\frac{k}{28.87})^2}$.

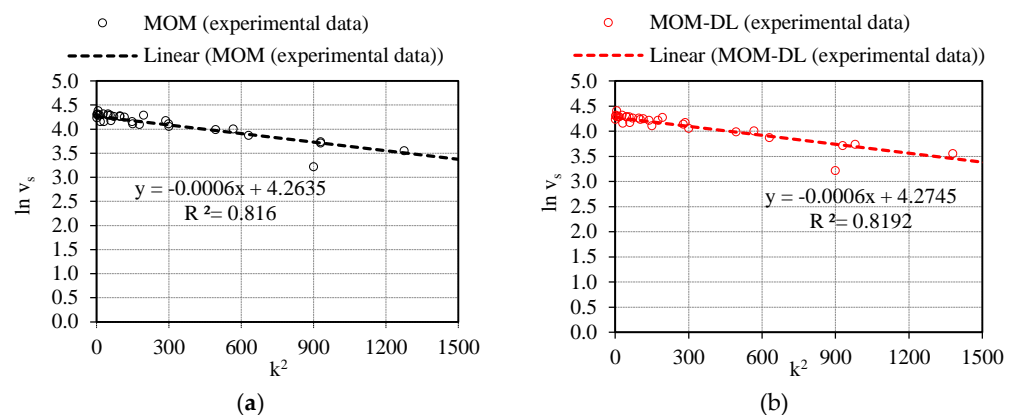


Figure 14. Speed-density scatter plots: (a) MOM method; (b) MOM-DL method.

Figures 15–17 illustrate the calibrated May’s traffic relationships $v_s = v_s(q)$, $v_s = v_s(k)$, $q = q(k)$ with the overlapping of the values of the experimental traffic variables measured by the proposed automated counting method based on deep learning (MOM-DL) in comparison with those measured by the conventional one (MOM). The two methods are in good agreement with each other but with the MOM-DL method, as stated above, some false positives (FP) and false negatives (FN) can be found in the detection step because of camera vibrations or adverse environmental conditions. In any case, it can be noted that the values of the traffic variables estimated with the proposed method (MOM-DL) are close to those of the traditional one (MOM). As a matter of fact, with respect to MOM, the MOM-DL technique can lead to lower values of some traffic flow variables such as the free-flow speed v_f , the capacity c , the critical speed v_c , and the critical density k_c , though the maximum percentage variation is relatively modest (2.3%, as shown in Table 2).

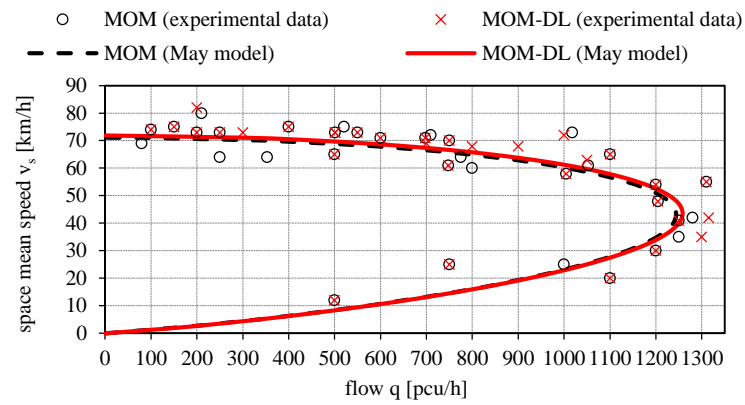


Figure 15. Pairs $(q; v_s)$ estimated for a lane of the SS640 highway (from section A to section B of Figure 7) and calibrated May traffic model.

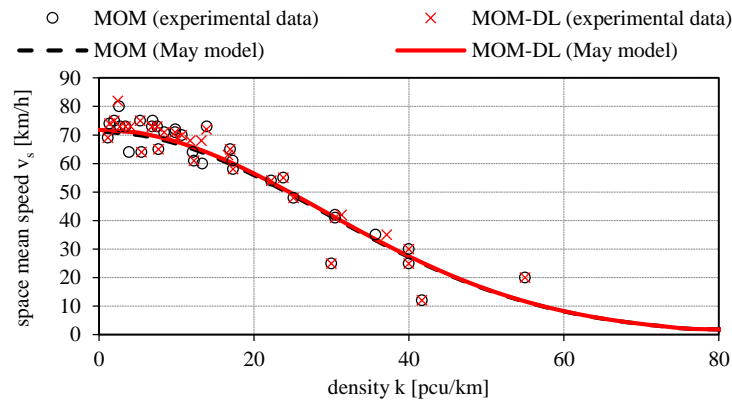


Figure 16. Pairs $(k; v_s)$ estimated for a lane of the SS640 highway (from section A to section B of Figure 7) and calibrated May traffic model.

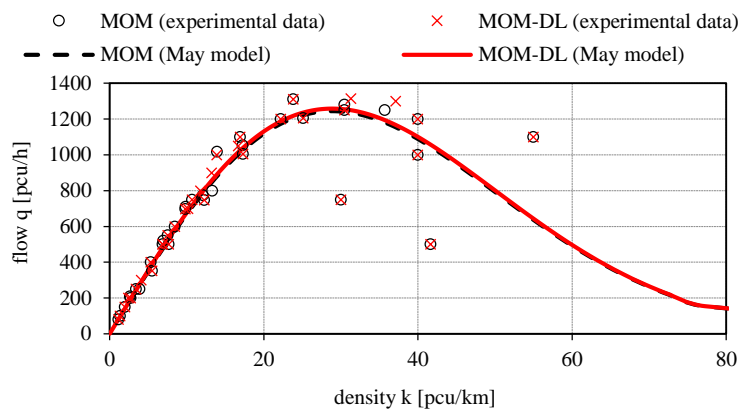


Figure 17. Pairs $(k; q)$ estimated for a lane of the SS640 highway (from section A to section B of Figure 7) and calibrated May traffic model.

Table 2. Comparison among the values of the traffic flow variables obtained with MOM-DL and MOM methods.

Method	Free-Flow Speed v_f [km/h]	Lane Capacity c [pcu/h]	Critical Speed v_c [km/h]	Critical Density k_c [pcu/km]
MOM	71.06	1244.15	43.00	28.87
MOM-DL	71.84	1257.80	44.00	28.87
Δ [%]	1.1	1.1	2.3	0.0

6. Conclusions

The evaluation of traffic flow variables is a key element in highway planning and design phases, as well as in traffic control strategies of existing infrastructures. Traffic data can be collected by numerous techniques, including the fixed spot measurement and the “probe vehicle data” systems. In addition, the moving observer method (MOM) developed by Wardrop can be used as a traffic data acquirement method. As is well known, artificial intelligence (AI) and deep learning (DL) are commonly used in numerous applications of real-life areas including vehicle detection and recognition.

This article presents a novel automatic traffic data acquirement system founded on MOM, deep learning and the YOLOv3 algorithm for the estimation of macroscopic traffic variables. The proposed method, called MOM-DL, allows the measurement of the flow rate (q), the space mean speed (v_s) and the density (k) in case of stationary and homogeneous traffic flow conditions. Being an automated method, MOM-DL does not necessitate human observers with consequent significant advantages in terms of user safety improvements and costs reductions. MOM-DL was subjected to strong verification, calibration and validation by means of real-world traffic datasets. Experiments have been conducted along a segment of an Italian highway (SS640) having a length of 1.1 km and uninterrupted flow conditions.

In accordance with the moving observer method, the survey vehicle has been driven numerous times in the same direction and in the opposite direction with respect to the traffic stream of flow q that must be measured. The test vehicle speed has been fixed in the range 60 ± 5 km/h depending on the traffic state condition. For each of the 35 trips, the vehicles belonging to the flow of interest have been recorded by means of a calibrated video camera installed in the survey vehicle.

The accuracy, loss and precision values obtained during the neural network training process prove that the proposed training model detects vehicles with high accuracy. The traffic flow variables have been calculated using the Wardrop relationships $q = q(x, y, t_a, t_w)$, $v_s = v_s(x, y, t_a, t_w)$, $k = k(x, y, t_a, t_w)$. Then, the empirical traffic data, in terms of pairs ($q; v_s$), ($k; v_s$), ($k; q$) were employed for the comparison between MOM-DL and MOM. The results of the research demonstrate that the two methods are in good agreement with each other, although some false positives and false negatives may be found in the detection step mainly due to camera vibrations.

The proposed method is marked by a reliable and rapid analytical approach as well as by high accuracy in the measurements of traffic variables. In fact, the research shows that the values of the macroscopic traffic variables estimated by means of the May traffic flow model together with the proposed method (MOM-DL technique) are very close to those obtained by the traditional one (MOM technique), with the maximum percentage variation being less than 3%.

Author Contributions: All authors substantially contributed to this work. Conceptualization, M.G.; methodology, M.G.; software, M.G. and G.P.; validation, M.G. and G.P.; investigation, M.G. and G.P.; writing—original draft preparation, M.G.; writing—review and editing, M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Projects of National Interest—PRIN 2017 “Stone pavements. History, conservation, valorisation and design” (20174JW7ZL) financed by the Ministry of Education, University and Research (MIUR) of the Italian Government.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors wish to express their gratitude to the anonymous reviewers for the comments and suggestions provided, which have been very helpful in improving the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Juan, PR, USA, 17–19 June 1997; IEEE: Piscataway, NJ, USA, 2016; pp. 770–778.
2. Li, Y.; Guo, J.; Guo, X.; Liu, K.; Zhao, W.; Luo, Y.; Wang, Z. A Novel Target Detection Method of the Unmanned Surface Vehicle under All-Weather Conditions with an Improved YOLOV3. *Sensors* **2020**, *20*, 4885. [CrossRef]
3. Chen, X.Z.; Chang, C.M.; Yu, C.W.; Chen, Y.L. A Real-Time Vehicle Detection System under Various Bad Weather Conditions Based on a Deep Learning Model without Retraining. *Sensors* **2020**, *20*, 5731. [CrossRef] [PubMed]
4. Goma, A.; Abdelwahab, M.M.; Abo-Zahhad, M.; Minematsu, T. Taniguchi RI. Robust Vehicle Detection and Counting Algorithm Employing a Convolution Neural Network and Optical Flow. *Sensors* **2019**, *19*, 4588. [CrossRef] [PubMed]
5. Biswas, D.; Su, H.; Wang, C.; Blankenship, J.; Stevanovic, A. An Automatic Car Counting System Using OverFeat Framework. *Sensors* **2017**, *17*, 1535. [CrossRef] [PubMed]
6. Wardrop, J.G.; Charlesworth, G. A Method of Estimating Speed and Flow of Traffic from a Moving Vehicle. *Proc. Inst. Civ. Eng.* **1954**, *3*, 158–171. [CrossRef]
7. Bi, F.; Yang, J. Target Detection System Design and FPGA Implementation Based on YOLO v2 Algorithm, International Conference on Imaging, Signal Processing and Communication. *ICISPC* **2019**, 10–14.
8. Lechgar, H.; Bekkar, H.; Rhinane, H. Detection of cities vehicle fleet using YOLO V2 and aerial images, International Archives of the Photogrammetry. *Remote Sens. Spat. Inf. Sci.-ISPRS Arch.* **2019**, *42*, 121–126.
9. Elgendy, M. *Deep Learning for Vision Systems*; Simon and Schuster: New York, NY, USA, 2020.
10. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. 2018. Available online: <https://arxiv.org/pdf/1804.02767v1.pdf> (accessed on 1 August 2021).
11. Siyal, M.Y.; Fathy, M. Image processing techniques for real-time qualitative road traffic data analysis. *Real-Time Imaging* **1999**, *54*, 271–278. [CrossRef]
12. Wang, L.; Yang, S.; Yang, S. Automatic thyroid nodule recognition and diagnosis in ultrasound imaging with the YOLOv2 neural network. *World J. Surg. Oncol.* **2019**, *17*, 12. [CrossRef]
13. Pan, Q.; Guo, Y.; Wang, Z. A scene classification algorithm of visual robot based on Tiny Yolo v2. In Proceedings of the 2019 Chinese Control Conference (CCC), Guangzhou, China, 27–30 July 2019; pp. 8544–8549.
14. Ge, L.; Dan, D.; Li, H. An accurate and robust monitoring method of full-bridge traffic load distribution based on YOLO-v3 machine vision. *Struct. Control. Health Monit.* **2020**, *27*, e2636. [CrossRef]
15. Yolo v3 of Yolo Series. Available online: <https://blog.csdn.net/leviopku/article/details/82660381> (accessed on 1 August 2021). (In Chinese)
16. Jin, Z.; Zheng, Y. Research on application of improved YOLO V3 algorithm in road target detection. *J. Phys. Conf. Ser.* **2020**, *1654*, 012060.
17. Dewi, C.; Chen, R.-C.; Yu, H. Weight analysis for various prohibitory sign detection and recognition using deep learning. *Multimed. Tools Appl.* **2020**, *79*, 32897–32915. [CrossRef]
18. Nie, M.; Wang, C. Pavement Crack Detection based on yolo v3. In Proceedings of the 2019 2nd International Conference on Safety Produce Informatization (IICSPI), Chongqing, China, 28–30 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 327–330.
19. Hui, J. Real-Time Object Detection with YOLO, YOLOv2 and Now YOLOv3. 2018. Available online: <https://jonathan-hui.medium.com/real-time-object-detection-with-yolo-yolov2-28b1b93e2088> (accessed on 1 August 2021).
20. Kang, H.; Chen, C. Fast implementation of real-time fruit detection in apple orchards using deep learning. *Comput. Electron. Agric.* **2020**, *168*, 105108. [CrossRef]
21. Guerrieri, M.; Parla, G.; Mauro, R. Traffic Flow Variables Estimation: An Automated Procedure Based on Moving Observer Method. Potential Application for Autonomous Vehicles. *Transp. Telecommun.* **2019**, *20*, 205–214. [CrossRef]
22. Bennett, T.H. A further procedure for estimating speed distribution parameters in uni-directional traffic streams using the moving observer method. *Transpn Res.* **1977**, *11*, 205–207. [CrossRef]

23. Duncan, N.C. *A Method of Estimating the Distribution of Speeds of Cars on Motorways*; TRRL, LR 598; Transport and Road Research Laboratory: Crowthorne, Berkshire, 1973.
24. Hewitt, R.H.; Chambers, J.B. Graphical solution of moving observer surveys, *The Highway Engineer. J. Inst Highw. Eng.* **1973**, *1*, 12–16.
25. Lee, B.H.; Brocklebank, P.J. *Speed-Flow-Geometry Relationships for Rural Single Carriageway Roads*; TRRL Contractor Report 319; Transport Research Laboratory: Crowthorne, UK, 1993.
26. Du, X.; Sun, C.; Zheng, Y.; Feng, X.; Li, N. Evaluation of vehicle vibration comfort using deep learning. *Meas. J. Int. Meas. Confed.* **2021**, *173*, 108634. [[CrossRef](#)]
27. Zhang, Z. A flexible new technique for camera calibration. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*; IEEE: Piscataway, NJ, USA, 2000; Volume 22, pp. 1330–1334.
28. Guerrieri, M.; Mauro, R.; Tollazzi, T. Turbo-Roundabout: Case Study of Driver Behavior and Kinematic Parameters of Light and Heavy Vehicles. *J. Transp. Eng. Part A Syst.* **2019**, *145*. [[CrossRef](#)]
29. Guerrieri, M.; Mauro, R.; Parla, G.; Tollazzi, T. Analysis of kinematic parameters and driver behavior at turbo roundabouts. *J. Transp. Eng. Part A Syst.* **2018**, *144*. [[CrossRef](#)]
30. Available online: <http://www.vision.caltech.edu/archive.html> (accessed on 1 August 2021).
31. Dollár, P.; Wojek, C.; Schiele, B.; Perona, P. *Pedestrian Detection: An Evaluation of the State of the ART*; IEEE: Piscataway, NJ, USA, 2012.
32. Dollár, P.; Wojek, C.; Schiele, B.; Perona, P. *Pedestrian Detection: A Benchmark*; CVPR: Miami, FL, USA, 2009.
33. Chen, Q.; Liu, L.; Han, R.; Qian, J.; Qi, D. Image identification method on high speed railway contact network based on YOLO v3 and SENet. In *Proceedings of the Chinese Control Conference (CCC)*, Guangzhou, China, 27–30 July 2019; pp. 8772–8777.
34. Raqib, A.; Ibrahim, W.H.W.; Sadullah, A.F.M.S. Estimating Travel Time of Arterial Road Using Car Chasing Method and Moving Observer Method. *J. Transp. Sci. Soc. Malays.* **2005**, *1*, 77–87.
35. Zhou, Y.; Pei, Y.; Li, Z.; Fang, L.; Zhao, Y.; Yi, W. Vehicle weight identification system for spatiotemporal load distribution on bridges based on non-contact machine vision technology and deep learning algorithms. *Measurement* **2020**, *159*. [[CrossRef](#)]
36. Gerlough, D.L.; Huber, M.J. *Traffic Flow Theory: A Monograph*; Special Report; TRB: Chennai, India, 1975; p. 165. Available online: http://tft.eng.usf.edu/docs/Traffic_Flow_Theory_Monograph_1975.pdf (accessed on 1 August 2021).
37. Guerrieri, M.; Mauro, R. *Macroscopic Traffic Flow Models, A Concise Introduction to Traffic Engineering*; Springer: Berlin/Heidelberg, Germany, 2021; ISBN 978-3-030-60723-4.
38. Cantisani, G.; Serrone, G.D.; Biagio, G.D. Calibration and validation of and results from a micro-simulation model to explore drivers' actual use of acceleration lanes. *Simul. Model. Pract. Theory* **2018**, *89*, 82–99. [[CrossRef](#)]