



Article

Interpolation-Based Framework for Generation of Ground Truth Data for Testing Lane Detection Algorithm for Automated Vehicle

Swapnil Waykole, Nirajan Shiwakoti * and Peter Stasinopoulos

School of Engineering, RMIT University, Melbourne, VIC 3000, Australia

* Correspondence: nirajan.shiwakoti@rmit.edu.au

Abstract: Automated vehicles, predicted to be fully electric in future, are expected to reduce road fatalities and road traffic emissions. The lane departure warning system, an important feature of automated vehicles, utilize lane detection and tracking algorithms. Researchers are constrained to test their lane detection algorithms because of the small publicly available datasets. Additionally, those datasets may not represent differences in road geometries, lane marking and other details unique to a particular geographic location. Existing methods to develop the ground truth datasets are time intensive. To address this gap, this study proposed a framework for an interpolation approach for quickly generating reliable ground truth data. The proposed method leverages the advantage of the existing manual and time-slice approaches. A detailed framework for the interpolation approach is presented and the performance of the approach is compared with the existing methods. Video datasets for performance evaluation were collected in Melbourne, Australia. The results show that the proposed approach outperformed four existing approaches with a reduction in time for generating ground truth data in the range from 4.8% to 87.4%. A reliable and quick method for generating ground truth data, as proposed in this study, will be valuable to researchers as they can use it to test and evaluate their lane detection and tracking algorithms.

Keywords: lane detection algorithm; electric vehicle; custom database; image processing; advanced driver assistance systems



Citation: Waykole, S.; Shiwakoti, N.; Stasinopoulos, P. Interpolation-Based Framework for Generation of Ground Truth Data for Testing Lane Detection Algorithm for Automated Vehicle. *World Electr. Veh. J.* **2023**, *14*, 48. <https://doi.org/10.3390/wevj14020048>

Academic Editors: Biao Yu, Linglong Lin and Jiajia Chen

Received: 14 December 2022

Revised: 2 February 2023

Accepted: 7 February 2023

Published: 9 February 2023



Copyright: © 2023 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

Traffic crashes are increasing and becoming one of the critical issues worldwide with the rapid development of expressways and the growth of motor vehicle numbers. In recent years, automated ground vehicles have emerged as an essential component of intelligent transportation systems (ITS). Further, automated vehicles, predicted to be fully electric in future, will reduce road traffic emissions. A fully automated vehicle will drive people to their destination without any shared control with the driver, including controlling safety-critical tasks. It is necessary to integrate computers, controls, communications, and different automation technologies in ITS in order to improve transportation safety, throughput, and efficiency, while lowering energy consumption and environmental impact [1]. The car communicates with the driver, the surroundings, and the infrastructure. These interactions are enhanced in intelligent cars by sensing, sharing information, and actuating different primary or secondary driving activities. The Advanced Driving Assistant Technology (ADAS) system is the foundation of ITS. The primary goal of ADAS systems is to enhance road safety, reduce traffic congestion, and increase driving comfort by increasing various ADAS features [2].

Nowadays, most automobile advances are driven by embedded technologies and software solutions that identify potentially unsafe driving scenarios. It is argued that ADAS may reduce human driving mistakes [3]. One key breakthrough in the automotive industry is the advent of electric cars, which presents various potentials for ADAS development, but

also many obstacles. Much research is being conducted in automated vehicles to develop enhanced ADAS to provide safety to drivers. Lane detection and tracking have been critical features of ADAS for safe driving and avoiding accidents.

Automated ground vehicle control systems can be divided into three components: environment prediction and planning; decision-making; and vehicle motion control. Vehicle motion control is separated into two categories: longitudinal motion control and lateral motion control. Longitudinal motion control is also known as longitudinal velocity tracking control, and it is a critical component of an intelligent driving control system [4]. Several control algorithms have been applied to longitudinal velocity tracking control. The longitudinal controller controls the vehicle speed, while the lateral controller controls the lateral offset and relative heading between the planned trajectories and the vehicle. The controller design challenge is addressed using a mathematical model that describes both lateral and longitudinal motions. The controller would ensure that the vehicle follows a speed profile while automatically regulating the vehicle speed to maintain a safe distance from the preceding vehicle [5]. A mixture of lane maintaining and overtaking in the left and right scenarios is utilized to test the controllers, which provide excellent tracking and comfort preservation [6,7].

Ground truth data in the context of computer vision involves a series of images, a set of labels on the images, and establishing a model for object recognition, including the number, location, and relationships of important characteristics, among other things (environmental factors). Depending on the complexity of the challenge, the labels are placed manually or automatically using image analysis.

The lane departure warning system (LDWS) utilizes lane detection and tracking algorithms and is now an important feature of ADAS or automated vehicles. The availability of high-quality datasets is important for developing new computer vision algorithms. In the literature, lane detection algorithms are usually classified under three categories: the model-based approach; a features-based approach; and a learning-based approach [8]. Learning-based approaches, on the one hand, require a large amount of ground truth data during their training process, whereas high-quality data are required for a detailed and equal assessment and comparison of other methods. Datasets and benchmark evaluations, which allow us to quantify progress in lane detection and tracking systems, have stimulated much innovation in the computer vision and machine learning fields. Some data types, such as camera images or depth images in indoor scenarios, are relatively simple to obtain with high precision. Low-cost sensors that can quickly produce a large amount of data are available for such image modalities. Manual user annotation can be used to obtain other ground truth data, such as object annotations or multiclass pixel-wise image annotations [8].

There are many large-scale datasets available for the performance evaluation of an algorithm of optical flow and object detection [9]. There are very few extensive datasets available for lane detection and tracking algorithms with a high-quality dataset of outdoor scenes, especially in the case of degraded lane markings. High-precision dense videos and huge annotated images are required to test an algorithm in ADAS. One way to obtain a large high-precision dataset is the ground truth dataset, which can be collected by mounting a camera and other equipment at the front of the vehicle. This is a time-consuming and costly procedure unless the ground truth data can be generated in a reasonable time. Other types of ground truth data, such as outdoor depth images and optical flow, are also rare to pull out. Stereo cameras and laser scanners may provide depth information, but they are either inaccurate or provide very sparse data.

Objective and Scope of the Study

This study aims to develop high-precision ground truth data based on an interpolation approach, which takes advantage of the existing manual and time-slice approaches. Researchers are constrained in testing their lane detection algorithms because of the low number of publicly available datasets. Additionally, those datasets may not represent differ-

ences in road geometries, lane marking, and other details unique to a particular geographic location. Researchers have developed a few procedures for collecting ground truth (custom) datasets, but it is time-consuming to generate high-precision data to test the algorithm. Therefore, to address this knowledge gap, this study develops an interpolation approach to quickly generate high-precision data for testing and evaluating the performance of lane detection and tracking algorithm.

This study provides methodological advancement in terms of quickly generating ground truth data (i.e., data with annotations added to them), which researchers can then use for testing lane detection algorithms. Without ground truth annotation, lane detection algorithms cannot be trained and tested. At the present time, researchers are constrained in testing their lane detection algorithms because of the low number of publicly available datasets. Further, the generation of ground truth annotation is mainly limited to a time-consuming and less reliable manual approach. In the manual approach, annotation depends on the user's skill or judgement; thus, the performance of lane annotation varies from person to person. For example, if two people with varied experience perform the manual approach, the annotation result may differ. This is one of the limitations we overcome with our proposed approach, besides the reduction in time for annotation.

Since ground truth annotation is a crucial step before testing lane detection algorithms, any method that can overcome the disadvantages of the time-consuming task of manual annotation is a methodological advancement. With our proposed semiautomatic interpolation-based framework for generating ground truth annotation, researchers can quickly make a reliable dataset to be ready for training and testing lane detection algorithms. Interpolation is a method by which related known values are used to estimate an unknown value or set of values. We have defined finite number rows and identified the unknown value through an interpolation approach to connect a smooth line or curve on the images by connecting these points. Our proposed approach allows researchers to quickly generate a variety of reliable ground truth datasets that they can deploy to test their lane detection algorithms for different road geometries, lane marking, and other details unique to a particular geographic location. Video datasets for performance evaluation of our proposed approach are collected in Melbourne, Australia. The performance of the proposed approach is tested by comparing results with other existing approaches (including manual annotation as a benchmark) to demonstrate the superiority of our proposed approach.

The structure of the paper is organized as follows. Section 2 presents a literature review on related works for the generation of ground truth data, a comparison of existing datasets, and approaches to develop ground truth data. It is followed by Section 3, which explains the data collection procedure adopted for this study and the proposed interpolation approach framework. Section 4 discusses the proposed approach's advantage over other existing approaches. Finally, conclusions and recommendations for future works are presented.

2. Literature Review

The following subsections discuss related work on generating ground truth or custom data, and the comprehensive analysis of available datasets. Data used for lane detection and tracking and the evaluation of algorithms are explained, along with existing approaches to develop ground truth data.

2.1. Related Work to a Generation of Ground Truth Data

Veit et al. [10] used the ROMA dataset to test a number of feature extractors for road markings. With camera details and ground truth for the lane markers, the ROMA dataset offer high-quality color images. The total length of its image sequence is below 20 s. In addition, it appears that the dataset is compiled from a set of random images that would make it inappropriate to test lane marking that combines algorithms with lane tracking. Nevertheless, datasets are also available that contain image sequences for testing. For instance, while driving on local city roads, Leibe et al. [11], Wanf et al. [12], and Brostow et al. [13] generated image sequences with different scenarios. Although the recorded image content

of the dataset had a fluctuation in illumination, different lane marking and surfaces, and the neighboring vehicles, the overall length of the image content is less than 10 min. In addition, ground truth files are not provided, which is this dataset's main drawback. While the dataset generated by Aly et al. [14] contains a ground truth file, scenes are not included in sequences of the image on the recorded videos on the highway. The PESTS2001 dataset [15] included image sequences of a highway, whereas the dataset from Sivaraman et al. [16] generated image sequences of a street and highway. Unfortunately, both datasets have compromised image quality by applying lossy compressions such as JPEG and MPEG-4. Finally, the datasets by Santos et al. [17], Lim et al. [18], and EISATS [19], provide a comprehensive collection of images that would be ideal for lane marking testing. The images in [17–19] display variations in the types of illumination, traffic flow conditions, road texture, and lane marking, that reflect the conditions. In addition, in these datasets, the total length of the sequence is more than 10 min, and even the camera parameters that are required for generating camera models are also given. These datasets are applicable for different applications, except for lane detection. Consequently, files containing ground truth data are unavailable, so objective analysis may not be obtained. In addition, testing algorithms on the dataset listed in [16–18] appear to measure their finding based on visual inspection, and may be biased. A dataset containing a large amount of image data along with ground truth is required to make a systematic, objective evaluation.

2.2. Comprehensive Analysis of the Available Dataset

The development of lane detection and tracking for ADAS is of great interest to the automotive society. However, access to resources is extremely restricted because of the significant commercial involvement in this area of research. Data that are used for assessment and research is one of the most valuable tools. Compared to areas of study such as face detection or the identification of optical characters (OCR), where labeled and structured datasets are often available for training and testing, there appear to be no such tools for lane detection and tracking. The issue caused by the absence of common datasets is that it is difficult to currently compare available lane detection algorithms easily. Additionally, the lack of common testing data make verifying other implementation results difficult. In this scenario, one of the viable solutions is to enforce each algorithm and its performance verification. However, the sheer quantity of algorithms that have already been published in journals and conferences makes this challenge difficult to achieve. Table 1 shows a comparison of different dataset features available in the existing literature for lane detection, including CU Lane [20], Caltech [21], NEXET [22], DIML [23], KITTI [24], TuSimple [25], UAH [26], and BDD100K [27].

Table 1. Comparison among different lane datasets.

Sr. No	CU Lane [20]	Caltech [21]	NEXET [22]	DIML [23]	KITTI [24]	TuSimple [25]	UAH [26]	BDD100K [27]
Sequences	More than 55 hrs videos	4 clips	Not available	Not available	22 sequences	7000 sequences	500 min videos	100,000 videos
Images	133,235	1225	50,000	470	14,999	140,000	N/A	120,000,000
Multiple cities	-	-	✓	✓	-	✓	✓	-
Multiple weathers	-	✓	✓	✓	-	-	✓	✓
Multiple times of day	✓	✓	✓	✓	-	-	✓	✓
Multiple scene types	✓	✓	✓	✓	✓	✓	✓	✓
Multiple cameras	✓	-	✓	✓	✓	-	-	✓
Multiple street	-	-	✓	✓	✓	-	✓	✓
Labelled streets	-	✓	✓	✓	-	✓	✓	✓

Note: ✓: Item under "Sr. No" column is included, and -: Item under "Sr. No" is excluded.

Researchers have used improvised datasets in terms of visual content to explore multitask learning for a self-driving car. In earlier decades, researchers were constrained because of a small set of datasets available for research. However, recently, Berkeley Deep Drive, University of California, Berkeley, developed the BDD100K dataset, consisting of 100K driving videos with diverse scene types for researchers. A dataset for structure has been covered in BDD100K according to different types of lanes, such as lane category, lane direction, lane continuity, and drivable area. In the lane category, the following types of lane markings have been covered with annotation: road curb, double white lanes, double yellow lanes, double other, single white lanes, single yellow lines, and other single and crosswalks. Additionally, parallel lanes and vertical lanes are annotated in the lane direction. Annotated images with full and dashed lanes are covered in lane continuity, and derivable and alternative types of lanes come under the drivable area.

Based on the review of studies on lane detection and tracking [8,9], and the summary of datasets presented in Table 1, it can be observed that there are limited datasets in the literature that researchers have used to test lane detection and tracking algorithms. It was also observed that there are a lack of specific datasets from Australian roads.

In future, more datasets may be available for researchers as this field continues to grow, especially with the development of fully automated electric vehicles. The verification of the performance of the algorithms for lane detection and tracking system is carried out based on the ground truth dataset. Ground truth data require a collection of images and a set of labels on the image in the context of computer vision and models for object recognition. The number, position, and relationships of key characteristics need to be included. Labels are added to the image either automatically or manually, depending on the complexity of the problem. The label set includes interesting points, the corners of lane boundaries and descriptors of features. Using a range of machine learning methods, a model can be trained, and the detected characteristics can be fed into a classifier at run time to calculate the correspondence between detected characteristics and modeled features.

2.3. Approaches for Developing Ground Truth Data

There are four popular approaches to developing ground truth data, as briefly described below:

(1) Manual approach

This method is transparent, straightforward, and often used to generate accurate ground truth. It is conducted as follows:

- The user can manually annotate the lane markings of the left lane's boundary at various points in an image.
- Same steps are repeated for the right lane boundary.
- All the steps are repeated for the video clip.
- Finally, the ground truth is generated and saved in the file.

This approach, however, has two issues: it is a slow procedure, and it requires proper details to annotate a single curve in an image, which can take up a lot of time. Gaps between two dashed lines are not annotated, because the lane boundary is difficult to estimate in these areas [28].

(2) Time-slice approach

By stacking an empty image with a row of pixels from frames in the video collection, a time-sliced (TS) image is produced. To further explain, each frame of a video set with F frames can be considered as an image with $M \times N$ dimensions [29].

(3) E-NET approach

In a study by [30], 168 images with a resolution of 1024×1280 pixels and normal RGB channels are included in the complete dataset. Around five to seven mast cells make up each image. About 40% of all cells have connected boundaries and appear to be near to one another. Images were divided into 240×560 tile segments, in accordance with the size of the neural networks. An expert manually segmented the dataset using the ground truth.

The complete dataset was divided into test and validation parts at a ratio of 74:51, with the samples being randomly shuffled [30].

(4) Automated test approach

The automatic data-labeling method for creating crack ground truths (GTs) within concrete images is presented in this approach [31]. The primary technique entails producing first-round GTs, pretraining a model based on deep learning, and producing second-round GTs. A learning-based crack detection model can be trained in a self-supervised way using the generated second-round GTs of the training data. After being retrained using the second set of GTs, the deep learning-based model that was previously trained is successful at detecting cracks. This method's primary purpose is the suggestion of an automated GT generating approach for pixel-level fracture detection model training [31].

3. Proposed Framework

In this work, we have proposed a framework for generating a ground truth dataset that can be used for testing lane detection algorithm. Our proposed approach leverages the advantages of manual and time-slice approaches through the interpolation approach. While the philosophy of the interpolation approach is applicable to all types of roads, we have used linear interpolation for straight roads and cubic spline interpolation for curved roads. The linear interpolation is the straight line connecting the two known positions, shown in Figure 1a. The equation of slopes gives the value of y along the straight line for x in the interval:

$$\frac{y^* - y_1^*}{y_2^* - y_1^*} = \frac{x^* - x_1^*}{x_2^* - x_1^*} \quad (1)$$

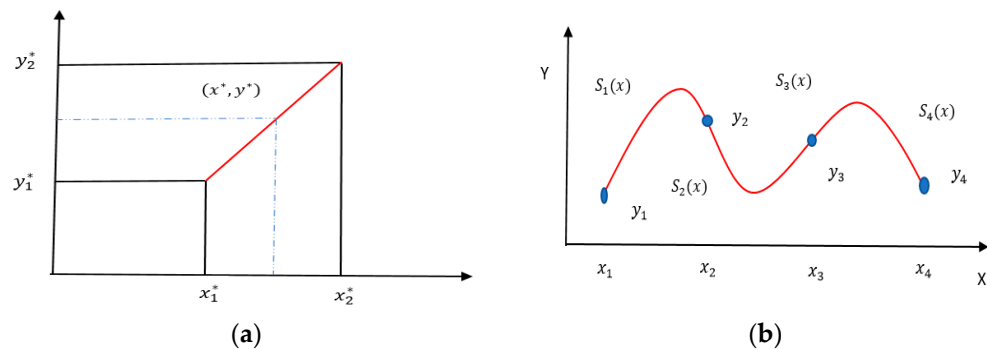


Figure 1. The concept of linear interpolation (a) and cubic spline interpolation (b).

y^* = linear interpolation values.

x^* = independent variable.

x_1^* & x_2^* = values of the function at one point.

y_1^* & y_2^* = values of the function at another point.

Cubic spline interpolation is a mathematical approach for creating new points within the boundary of a given collection of points, shown in Figure 1b. These new points are function values of an interpolation function composed of numerous cubic piecewise polynomials. The cubic spline is the function $S(x)$ with given properties $S(x)[x_i, x_i + 1] = S_i(x)$.

On the basis of the interpolation principle, the ground truth dataset was developed based on the videos obtained using monocular camera image sequences of traffic scenes in Melbourne, Australia. We have used the Driving Scenario Designer app (available in MATLAB R2021a) and MATLAB tools to annotate images.

It is difficult to estimate lane boundaries in the gaps, and to address this challenge, we have selected user-annotated points to specify a maximum number of rows close to each other. Then, we created cubic interpolation in these annotated points to get a smooth curve because lane markings are not straight everywhere (for straight roads, we used linear interpolation), so the controller can predict the lane boundaries based on the optimization. In this way, we can generate a dataset for lane detection and tracking algorithms in arbitrary

scenarios with different traffic flow conditions. Finally, a comparison of the performance of the proposed approach with other existing methods is conducted.

3.1. The Requirement of Ground Truth Dataset

The image data and the metadata are required to formulate a full dataset. The image data must consist of a series of image sequences captured at a speed of over 20 km/h while driving. Additionally, the time between the end of one sequence and the start of second recording is suggested to be a minimum of 2 min. This prevents the image data from being a collection of random images in the dataset. The criteria for the dataset from captured images are the following:

1. Duration of sequences: each recorded image sequence must have a length of at least 20 s. This approach will help test algorithms to use a frame-to-frame detection consistency.
2. Color: recorded image must be in 24-bit RGB format.
3. Uncompressed storage: recorded image sequence must store in a lossless format. There are two reasons for putting this provision in place. Some researchers have used real-time systems for research, in this case, real-time systems analyze files directly from the camera. So, the dataset's images must represent the camera's performance. Lossy compression creates artefacts of the compression and damage image quality.
4. Image format: to provide sufficient detail of the road surface, image must be at least (480 * 640) resolution.
5. Camera parameters: information about the intrinsic and extrinsic camera's parameters should be provided for camera models to be recreated. Focal length, the field of view, and image format, are among the intrinsic parameters. The extrinsic parameters include the ground plane's yaw, pitch, roll, and the height of the optical center of the sensor.
6. Lane marking: lanes are on straight and curving paths, with solid and dashed lane markers. They are the most common lane markers on roads in Australia. They must be present in the image details. The image details must also include images where lane markers are missing, to test for false positives.
7. Type of road: captured image data must include scenes from both streets and motorways.
8. Illumination effect: scenes in images that have various lighting changes, such as a vehicle driving via tunnels, under bridges, under shadows, etc., during the day time, night-time, and cloudy conditions. Data should be recorded to reflect variation in the levels of ambient light.
9. Weather condition: recorded image must include the weather condition, such as dry, rain, or snow.
10. Traffic flow condition: image must give traffic flow situations, such as heavy traffic, modern traffic, and light traffic, etc.
11. The ground truth describes the lane boundaries of the road in curve form, and these requirements are implemented by the shape of the roads and the position of the curve identified on the road. Curves identified on the lane boundaries should be smooth, and they must exhibit no visible kinks in their lane boundaries.
12. Generated curve must be at the center of the lane markers: if there are double lane markers, it should be between two lines. The curve must start at the top of the Region of Interest (ROI) and extend to the bottom of the image, along the lane boundary, after specifying a ROI.
13. In the presence of either solid or dashed markers, the lane boundary curve must be given.

3.2. Data Collection

To generate the ground truth datasets, we collected video clips covering more than 120 km of distance, including day and night, and under different weather conditions in four suburbs (Bundoora, Heidelberg, Kew, and Brighton) of Melbourne, Australia. Additionally, it includes both unstructured (no clear lane markings) and structured roads (clear lane

markings). Some videos were taken during the day time in Bundoora, Heidelberg, and Kew, with structured and unstructured roads, while some videos were collected in Brighton during the night. The parameters of the lens of the monocular camera used for data collection are shown in Table 2. We used this camera system because the ultracompact set provides a range of CMOS image sensors. The camera allows some image processing features, such as gamma correction and color interpolation, to plot a smooth curve with connecting points. The parameters of the lens are shown in Table 2.

Table 2. Parameters of the lens used in the camera.

Attribute of Lens	Parameters of Lens
Name of the company	Optics
Model number	60-255
Focal length	5-55(Mm)
Aperture	F1.3-1.6 C
Sensor	Not available
Working distance	800-1000(Mm)
Lens design	Back surface

For creating a ground truth dataset, it includes the following steps:

Model design: the model explains the structure of the objects, such as intensity, count, and location, and the relationship between a group of scale-invariant feature transform (SIFT) features. The model should be properly adjusted to the problem and image knowledge in order to yield significant outcomes.

Training dataset: in order to work with the model, this set was collected and labelled, and includes both positive and negative images and different features. Negatives include images and attributes that are intended to generate false matches on the lanes [32].

Test set: several images are collected for testing against the training set to predict the accurate match for the model.

Classifier design: this is designed to meet the speed and accuracy of the application objectives, including data organization and model search optimizations.

Training and testing: this dataset was collected to verify a group of images against ground truth [33].

Table 3 illustrates various road traffic environment features captured in the video data, collected in Melbourne, while Figure 2 summarizes the overall data collection process to postprocessing and evaluation.

Table 3. Road traffic environment features captured in the video data collection.

Features	Clip 1	Clip 2	Clip 3	Clip 4	Clip 5
Weather condition	Cloudy	Heavy rain in the night-time	Sunny	Cloudy	Heavy rain in the day time
Road surface	Rough surface	Rough surface	Smooth	Smooth	Rough
Color of lane marking	White	White	White	White	White
Traffic situation	Modern	Modern	Light	Light	Light
Speed	60 km/h	68 km/h	70 km/h	64 km/h	66 km/h
Number of frames	14	17	21	25	11
Type of lane marking	Dash	Solid	Dash and solid	Dash and solid	Dash and solid
Timing	Day time	Night-time	Day time	Evening	Day
Location	Heidelberg	Brighton	Bundoora	Bundoora	Kew
Structured and unstructured roads	Yes	Yes	Yes	Yes	Yes
Clothoid roads	Yes	Yes	Yes	Yes	Yes

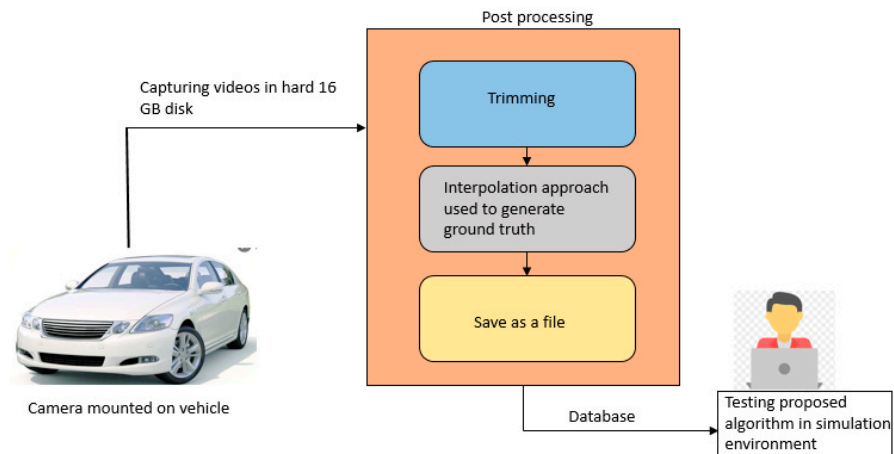


Figure 2. Summary of the procedure adopted to develop ground truth data.

3.3. Proposed Interpolation Approach

The proposed interpolation approach is a quick and effective technique for processing reliable ground truth data. One of the benefits of this method over the manual approach is that interpolation values give the spatiotemporal representation of the video clip; as a result, annotations can be carried out rapidly and with better accuracy.

We have divided this interpolation approach into three major steps.

Step 1:

In stage 1, we define a finite number of rows in which the boundary of the lane is annotated. These rows are shown in Figure 3. It is important to define at least three rows or more for producing a smooth curve of the lane boundaries for a good result. First, we started by defining only two rows, but noticed it was not of sufficient accuracy. So, we tested with three, or more than three rows (mostly three to four rows) using linear interpolation (for straight roads) or cubic spline interpolation (for curved roads).

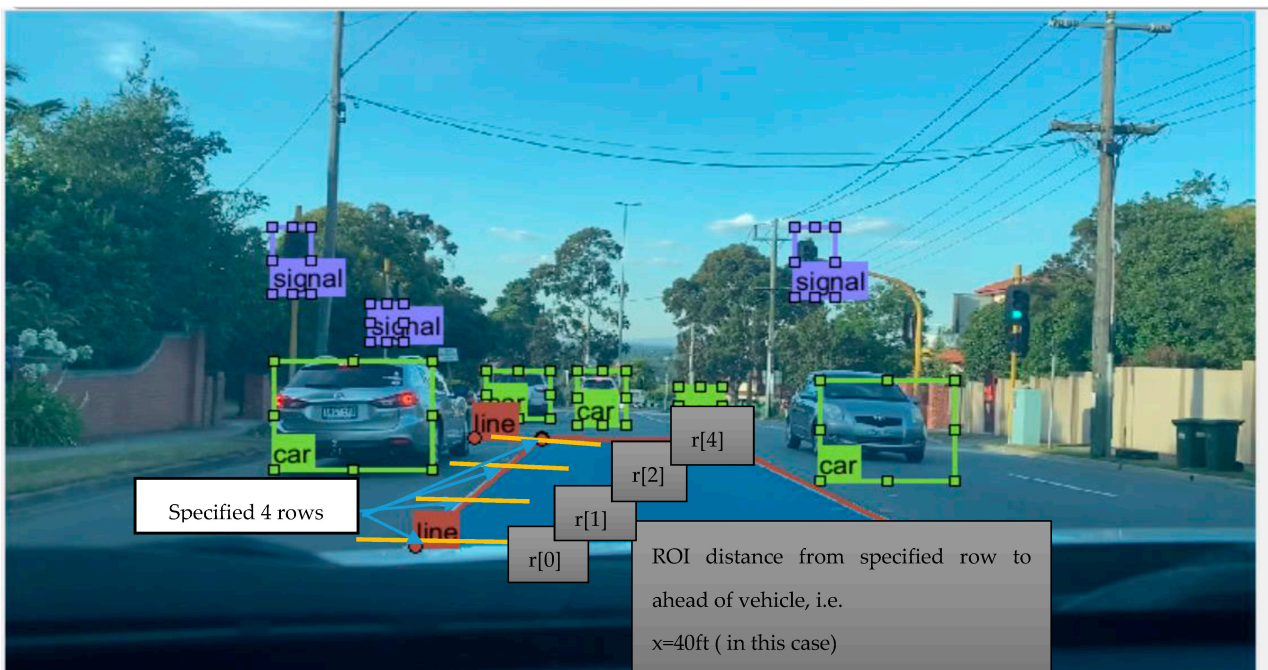


Figure 3. Defining a finite number of rows (four rows in this case) in the image.

It is to be noted that, in Figure 3, other markups, such as cars, signals, etc., are also shown. This is because we want to demonstrate (as a sample) how ground truth annotation for lane findings can be applied to a supervised lane detection and tracking algorithm, which also need to detect surrounding obstacles and objects to avoid a collision.

Step 2:

This stage is made up of two substages put within a frame. The specification of each substage is provided below.

1. Annotations: an interpolation image is generated by filling an empty image (without annotation) in the video clip with rows of image pixels. The sliced image, which is initially an empty image, is then generated with dimensions $F \times N$, where F is the video clip's frame count, and each image has dimension $M \times N$ in the video clip. The sliced image has dimensions of $F \times N$ as F copies of N pixels wide, unlike the images in the video clip. We have used 240×640 image size in this study. In the empty image, rows will be specified. The row corresponds to the sliced image row to simplify the nomenclature in this explanation, while the row is a row in the image in the video clip. Next, in the image, we specify the row from which pixels will be copied. Then, a single row of pixels from each image of the video clip is copied to a specific row in the empty image by using a loop. To elaborate, from the first image in the video clip (Figure 3), row = 300, 400, and 500. The first row of the sliced image is copied. Then, from the row = 300, 400, and 500 in time-sliced, the second image in the video clip is copied to the second-row image. Likewise, row = 300 from the F , the image will be copied to the F row in the time-sliced. The rows of pixels appear to be stacked in the time-sliced image since they are sequentially copied and positioned. The generated time-slice image is shown in Figure 4, where rows are copied from the image on the left to the sliced image on the right.

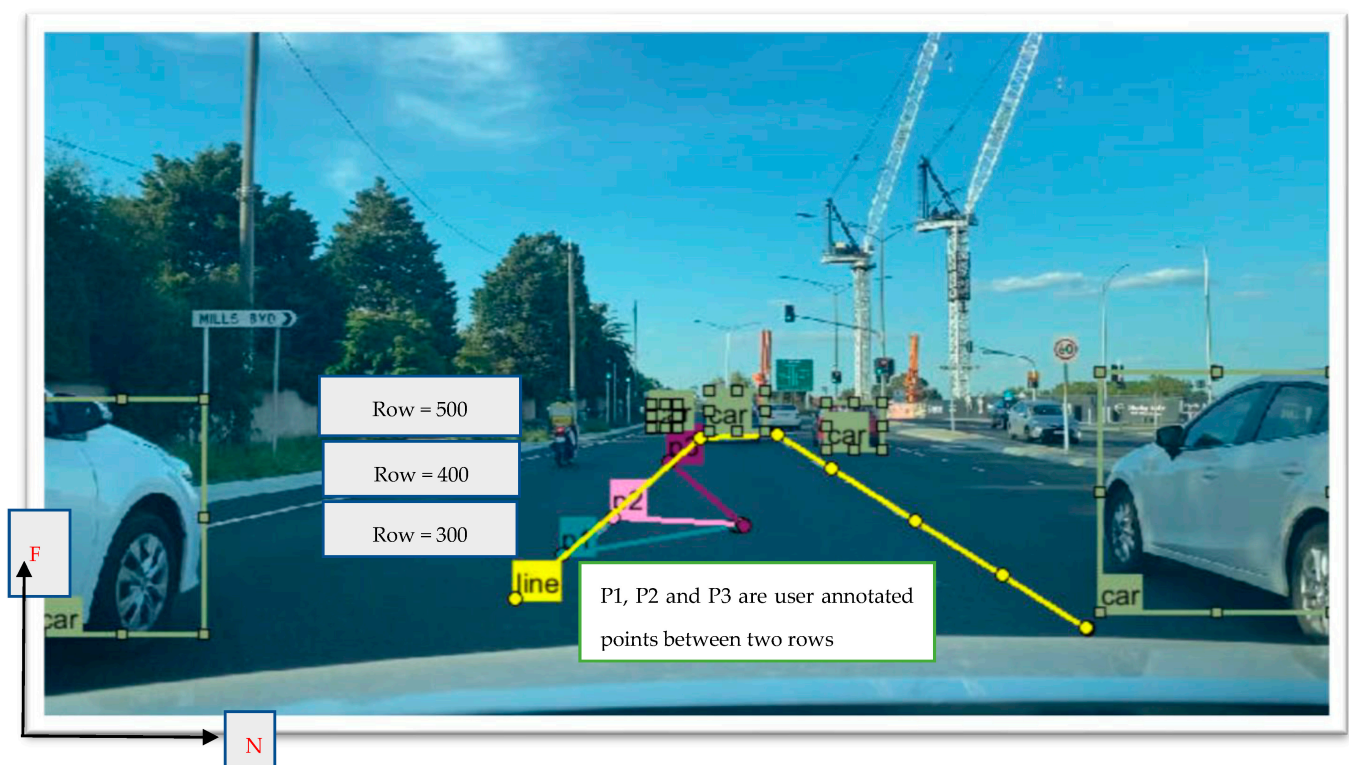


Figure 4. Connecting specified rows via interpolation to generate a smooth curve.

2. Interpolation: in this step, we first annotate a few points in the time-sliced image at the center of the left lane marker. Then, we perform curve fitting to link the points by interpolation (Figure 4). Instead of the nearest neighbor or linear interpolation, cubic spline interpolation is used here. It produces a smooth curve between the points, as shown in

Figure 4. In the interpolation image, curve fitting estimates lane marker positions between the annotated points (p_1 , p_2 , p_3 , along the yellow line), as shown in Figure 4.

Step 3:

At the end of stage 2, in each image of the video clip, we have lane boundary points in four rows. However, it will not create a smooth curve by simply connecting the four positions with straight lines. Therefore, again we use interpolation. However, this time, in each image, interpolation is performed across the rows, as shown in Figure 5.

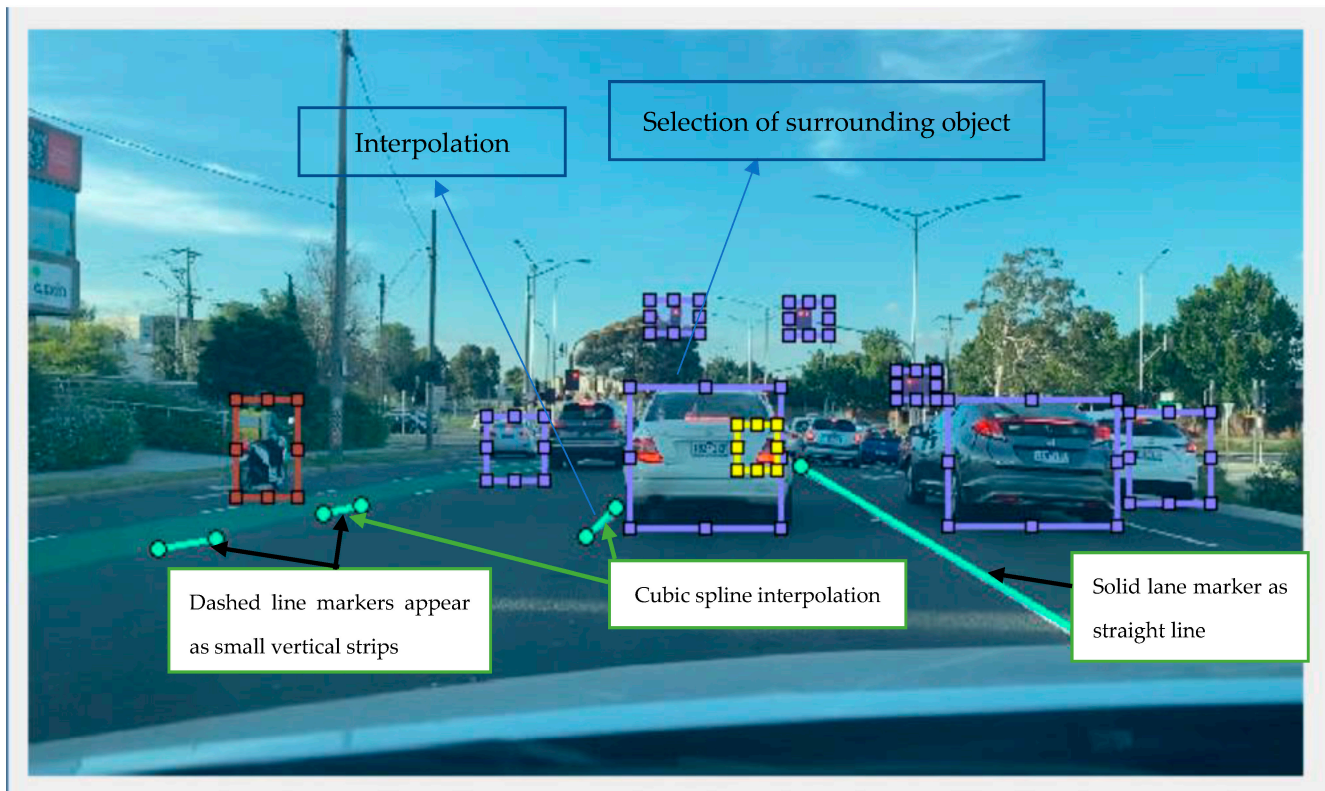


Figure 5. Generating ground truth annotation.

To elaborate, we have selected three points on the left lane boundary. The interpolation of the cubic spline binds the four rows and defines the position of the lane boundary on rows 300–500 at the same time, as shown in Figure 4. The cubic spline results in a smooth curve from $r[0]$ to $r[4]$ for the left lane boundary in the image.

Similarly, this procedure is carried out for the right lane boundary and then repeated for all the images in the video clip. Figure 4 displays the position on rows 300–500 for image 1 of the left and right lane boundaries. Furthermore, a sample interpolation image is shown in Figure 5 above this image, and two observations can be made:

1. Solid lane markers observed as a straight line;
2. Dashed lane markers observed small vertical strips.

Figure 6 summarizes the three steps involved in the interpolation approach. The XML file generated from the proposed interpolation approach is presented in the Appendix A (Figure A1).

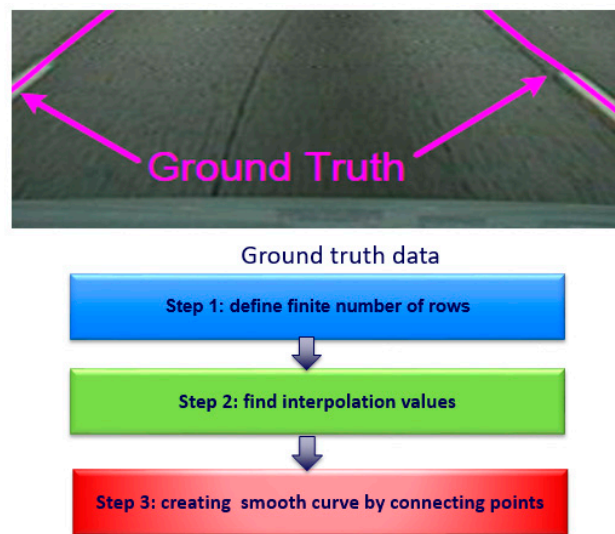


Figure 6. Summary of steps involved in the interpolation approach.

4. Discussion

The proposed approach focuses on interpolation to automate user-annotated points and, therefore, speed up the process of generating ground truth data. A reliable and quick method for generating ground truth data will be of immense importance to researchers as they can use it to test and evaluate their lane detection and tracking algorithms. We compared the performance (time taken to process and generate the ground truth data) of our proposed method with other existing methods: manual [28], time-slice [29], E-Net [30], and automated test [31], and the results are shown in Table 4. For comparison, we used different video clips with various duration, as shown in Table 4. All clips have solid and dashed lane markings. On average, from Table 4, it can be seen that our proposed interpolation approach took less time (5.78 min) as compared to other existing methods, such as the manual (46 min), time-slice (6.07 min), E-NET (6.35 min) and automated tests (6.97 min). This is a reduction of 87.4%, 4.8%, 8.9%, and 17.1%, respectively, in the processing time. When implementing lane marker classification in different weather conditions, the challenge is with ROI selection since the necessary image characteristics could be influenced by weather conditions, such as rain. While the proposed interpolation approach outperformed previous approaches in all environments, some extra time was needed for rainy conditions, compared to other normal days. For example, as compared to normal weather (sunny and dry: clip 3), it took around 16 s longer to extract the lane marker features and obtain interpolation values between defined rows in heavy rain at night-time (clip 2), and around 9 s longer in heavy rain during the day time (clip 5).

As interpolated values are obtained in our proposed approach, it has the capability to define the exact location of the lane boundaries. Further, the interpolation approach is not restricted to any specific lane marking, and road structure and weather conditions do not influence the ground truth dataset. The selection of an appropriate number of rows for interpolation may be a limitation, but can be overcome with the experience. At this stage of understanding, we believe that our proposed interpolation approach will provide a better result for testing the lane detection and tracking algorithm. This may explain why our approach is suitable for all types of lane marking. Table 4 shows the comparison of the interpolation approach with other approaches.

Table 4. Comparison of the proposed interpolation approach with other approaches.

Clip and Duration	Duration for Processing (mins)				
	Manual Approach	Interpolation Approach	E-NET	Automated Test Approach	Time-Slice Approach
Clip 1 [1 min]	47	5.00	6.00	8.00	6.01
Clip 2 [1.12 min]	41	4.56	5.00	7.32	5.00
Clip 3 [1.19 min]	53	7.00	7.30	6.53	7.09
Clip 4 [1.43 min]	44	6.37	6.70	7.00	8.00
Clip 5 [1.36 min]	45	6.00	6.78	6.04	7.04
Average	46	5.78	6.35	6.97	6.07
Standard Deviation	4.48	0.99	0.88	0.74	1.15

5. Conclusions

This study developed an interpolation approach for quickly generating reliable ground truth data. The proposed method takes advantage of the existing manual and time-slice approaches. The interpolation approach has been developed in three different steps: define a finite number of rows; find interpolation values; and create a smooth curve by connecting interpolation values. The system can be used to quickly produce large numbers of high-quality camera images, depth and optical flow videos, and textual annotations at the pixel level.

Furthermore, the proposed framework enables the creation of ground truth data for complex driving scenarios, which could be useful in developing lane detection and tracking systems for advanced driver assistance systems, or for testing algorithms for automated vehicles. Other applications of this proposed framework could be the determination of visual odometry, and the detection of motion models and scenes. Further, due to the semiautomatic nature of the proposed approach, the user can annotate each image by clicking a mouse button. Future work may explore combining data acquired from the front and rear cameras to determine the vehicle's position on the road, and identify lane detection and tracking.

Author Contributions: Conceptualization, investigation, data collection, methodology, writing—original draft preparation, S.W.; supervision, writing—review and editing, N.S.; supervision, P.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Acknowledgments: The first author would like to acknowledge the Government of India, Ministry of Social Justice and Empowerment, for providing a full scholarship to pursue Ph.D. study at RMIT University.

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

Appendix A

```

<Ground Truth>
  <Author>Swapnil Waykole </Author>
  <Institute>RMIT</Institute>
  <Date>20/01/2021 6:16:21 AM</Date>
  <ID>S1C1</ID>
  <CamNum>0</CamNum>
  <FrameCount>1376</FrameCount>
  <Annotation>
    <Fr ID="1">
      <Left>
        <X>236.609 235.220 234.831 ... 65.873 65.000</X>
        <Y>290.001 292.000 293.000 ... 419.000 420.000</Y>
      </Left>
      <Right>
        <X> ... </X>
        <Y> ... </Y>
      </Right>
    </Fr>
    <Fr ID="2"> ...
  
```

Figure A1. XML file generated from the proposed interpolation approach.

References

1. Eskandarian, A. *Handbook of Intelligent Vehicles*; Springer: London, UK, 2012.
2. Gagliardi, G.; Lupia, M.; Cario, G.; Casavola, A. Optimal H_∞ Control for Lateral Dynamics of Autonomous Vehicles. *Sensors* **2021**, *21*, 4072. [[CrossRef](#)] [[PubMed](#)]
3. Galvani, M. History and future of driver assistance. *IEEE Instrum. Meas. Mag.* **2019**, *22*, 11–16. [[CrossRef](#)]
4. Hang, P.; Chen, X.; Zhang, B.; Tang, T. Longitudinal Velocity Tracking Control of a 4WID Electric Vehicle. *IFAC-PapersOnLine* **2018**, *51*, 790–795. [[CrossRef](#)]
5. Gagliardi, G.; Casavola, A.; Toscano, S. Linear Parameter Varying Control Strategies for Combined Longitudinal and Lateral Dynamics of Autonomous Vehicles. In Proceedings of the 2022 European Control Conference (ECC), London, UK, 12–15 July 2022; pp. 181–186. [[CrossRef](#)]
6. Hima, S.; Lusseti, B.; Vanholme, B.; Glaser, S.; Mammari, S. Trajectory Tracking for Highly Automated Passenger Vehicles. *IFAC Proc. Vol.* **2011**, *44*, 12958–12963. [[CrossRef](#)]
7. Hima, S.; Glaser, S.; Chaibet, A.; Vanholme, B. Controller design for trajectory tracking of autonomous passenger vehicles. In Proceedings of the 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC), Washington, DC, USA, 5–7 October 2011; pp. 1459–1464. [[CrossRef](#)]
8. Waykole, S.; Shiwakoti, N.; Stasinopoulos, P. Review on Lane Detection and Tracking Algorithms of Advanced Driver Assistance System. *Sustainability* **2021**, *13*, 11417. [[CrossRef](#)]
9. Waykole, S.; Shiwakoti, N.; Stasinopoulos, P. Performance Evaluation of Lane Detection and Tracking Algorithm Based on Learning-Based Approach for Autonomous Vehicle. *Sustainability* **2022**, *14*, 12100. [[CrossRef](#)]
10. Veit, T.; Tarel, J.; Nicolle, P.; Charbonnier, P. Evaluation of Road Marking Feature Extraction. In Proceedings of the IEEE Conference on Intelligent Transportation Systems, Beijing, China, 12–15 October 2008; pp. 174–181.
11. Leibe, B.; Cornelis, N.; Cornelis, K.; Van Gool, L. Dynamic 3D Scene Analysis from a Moving Vehicle. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–8.
12. Wang, C.C. CMU/VASC Image Database. 2003. Available online: <http://vasc.ri.cmu.edu/idb/html/road/index.html> (accessed on 15 December 2022).

13. Brostow, G.J.; Shotton, J.; Fauqueur, J.; Cipolla, R. Segmentation and Recognition using Structure from Motion Point Clouds. In Proceedings of the European Conference on Computer Vision, Marseille, France, 12–18 October 2008; pp. 1–14.
14. Aly, M. Real time detection of lane markers in urban streets. In Proceedings of the IEEE Intelligent Vehicles Symposium, Eindhoven, The Netherlands, 4–6 June 2008; pp. 7–12.
15. Makris, D. PETS2001 Dataset. 2001. Available online: <http://www.cvg.cs.rdg.ac.uk/PETS2001/pets2001-dataset.html> (accessed on 15 December 2022).
16. Sivaraman, S.; Trivedi, M.M. A General Active-Learning Framework for On-Road Vehicle Recognition and Tracking. *IEEE Trans. Intell. Transp. Syst.* **2010**, *11*, 267–276.
17. Santos, V.; Almeida, J.; Gameiro, D.; Oliveira, M.; Pascoal, R.; Sabino, R.; Stein, P. ATLASCAR Technologies for a Computer Assisted Driving System on board a Common Automobile. In Proceedings of the IEEE Conference on Intelligent Transportation Systems, Funchal, Portugal, 19–22 September 2010; pp. 1421–1427.
18. Lim, K.H.; Cat, A.; Ngo, L.E.; Seng, K.P.; Ang, L.-M. UNMC-VIER Auto Vision Database. In Proceedings of the 2010 International Conference on Computer Applications and Industrial Electronics, Kuala Lumpur, Malaysia, 5–8 December 2010; pp. 650–654.
19. Multimedia Imaging Technology, Image Sequence Analysis Test Site (EISATS). Available online: <http://www.mi.auckland.ac.nz/EISATS/> (accessed on 18 November 2022).
20. Cu Lane Dataset. Available online: <https://xingangpan.github.io/projects/CULane.html> (accessed on 13 April 2020).
21. Gregory, G.; Holub, A.; Perona, P. Caltech-256 Object Category Dataset. California Institute of Technology. 2007. Available online: <https://resolver.caltech.edu/CaltechAUTHORS:CNS-TR-2007-001> (accessed on 15 October 2022).
22. Klein, I. NEXET—The Largest and Most Diverse Road Dataset in the World. 2007. Available online: <https://data.getnexar.com/blog/nexet-the-largest-and-most-diverse-road-dataset-in-the-world/> (accessed on 21 October 2022).
23. Lee, E. Digital Image Media Lab. Diml.yonsei.ac.kr. 2020. Available online: <http://diml.yonsei.ac.kr/dataset/> (accessed on 13 April 2020).
24. Cvlb.net. The KITTI Vision Benchmark Suite. Available online: <http://www.cvlib.net/datasets/kitti/> (accessed on 27 April 2020).
25. Tusimple/Tusimple-Benchmark. Available online: https://github.com/TuSimple/tusimple-benchmark/tree/master/doc/velocity_estimation (accessed on 15 April 2020).
26. Romera, E.; Luis, M.; Arroyo, L. Need Data for Driver Behavior Analysis? Presenting the Public UAH-Drive Set. In Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems, Rio de Janeiro, Brazil, 1–4 November 2016.
27. BDD100K Dataset. Available online: <https://mc.ai/bdd100k-dataset/> (accessed on 2 April 2020).
28. Coudray, N.; Karathandu, A.; Chamborn, S. Multi resolution approach for fine structure extraction—Application and validation on road images. In Proceedings of the Fifth International Conference on Computer Vision Theory and Applications, Angers, France, 17–21 May 2010; Volume 2.
29. Al-Sarraf, A.; Shin, B.S.; Xu, Z.; Klette, R. Ground Truth and Performance Evaluation of Lane Border Detection. In *Computer Vision and Graphics; ICCVG 2014. Lecture Notes in Computer Science, Volume 8671*; Chmielewski, L.J., Kozera, R., Shin, B.S., Wojciechowski, K., Eds.; Springer: Cham, Switzerland, 2014. [CrossRef]
30. Karimov, A.; Razumov, A.; Manbatchurina, R.; Simonova, K.; Donets, I.; Vlasova, A.; Khramtsova, Y.; Ushenin, K. Comparison of UNet, ENet, and BoxENet for Segmentation of Mast Cells in Scans of Histological Slices. In Proceedings of the 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), Novosibirsk, Russia, 21–27 October 2019; pp. 0544–0547. [CrossRef]
31. Chen, H.-C.; Li, Z.-T. Automated Ground Truth Generation for Learning-Based Crack Detection on Concrete Surfaces. *Appl. Sci.* **2021**, *11*, 10966. [CrossRef]
32. He, X.; Zemel, R.; Carreira-Perpin, M. Multiscale conditional random fields for image labeling. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 27 June–2 July 2004.
33. Shotton, J.; Winn, J.; Rother, C.; Criminisi, A. Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Graz, Austria, 7–13 May 2006.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.