IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 4 December 2022, accepted 18 December 2022, date of publication 26 December 2022, date of current version 2 January 2023.

Digital Object Identifier 10.1109/ACCESS.2022.3232127

RESEARCH ARTICLE

Learning Light Fields for Improved Lane Detection

MUHAMAD ZESHAN ALAM[1,2], SOUSSO KELOUWANI[2], JONATHAN BOISCLAIR[2], AND ALI AKREM AMAMOU[2]

[1]Department of Computer Science and Mathematics, University of Brandon, Brandon, MB R71 6A9, Canada
[2]Department of Mechanical Engineering, University of Quebec at Trois Rivieres, Trois Rivières, QC G8Z 4M3, Canada

Corresponding author: Muhamad Zeshan Alam (Muhammad.alam@uqtr.ca)

ABSTRACT Robust lane detection is imperative for the realization of intelligent transportation. Recently, vision-based systems that employ deep convolution neural networks (CNNs) for lane detection have made considerable progress. However, for better generalization under various road conditions learning-based methods require excessive training data, which becomes non-trivial in challenging conditions such as illumination variation, shadows, false lane lines, and worn lane markings, etc. In this paper, we propose a light field (LF) based lane detection method that utilizes the additional angular information for improved prediction and increased robustness. Two different LF representations are investigated to study the possibility of maximum performance improvement and minimal additional computation cost and data labeling efforts. Experimental results successfully demonstrate that the proposed approach improves the prediction of the lane line point coordinates and is significantly robust against the aforementioned adverse conditions.

INDEX TERMS Lane detection, light field imaging, convolutional neural networks, intelligent transportation.

I. INTRODUCTION

One of the major contributing factors of road traffic accidents (RTAs) can be attributed to human error, which can be caused by several factors, including but not limited to distractions, fatigue, and misbehavior [1]. To improve driving safety and minimize RTAs, ideally, human involvement needs to be minimized in road perception. For this purpose various advanced driver assistance systems (ADAS) particularly lane departure warning system (LDWS) and lane-keeping assistance system (LKAS) [2], [3] are developed to facilitate lane detection which is an integral part of the overall environmental perception.

The development of intelligent transportation, in general, and self-driving vehicles, in particular, demands a high level of accuracy and robustness for lane detection systems. To achieve level 5 autonomy, as defined by the international standard J3016 [4], the vehicles should be able to operate out of the so-called operational design domain. Instead of a carefully managed (usually urban) environment with lots of dedicated lane markings or infrastructure, it should be able to self-drive anywhere. Vehicles also must be able to detect road elements in all possible conditions. However, the frequently encountered scenarios such as illumination variations, shadows of trees and nearby vehicles covering the road, the appearance of false lines, and partially visible lane lines (due to occlusion and erosion over time) further increase the complexity of lane detection. In recent years vision-based lane detection systems are becoming increasingly popular. Automotive enterprises, such as Mobileye, BMW, and Tesla have developed their own vision-based lane detection and lane-keeping systems motivated by the following compelling factors: lane markings are painted on the road based on human visual perception, the decrease in the cost of high-end machine vision camera devices, and rapid advancements in deep convolution neural networks (CNNs) for vision-related tasks.

Deep CNNs obtain relevant features through training using multiple kernels layers. These kernels are updated as the algorithm is fed labeled data, converging by numerical optimization methods on the weights that best match the training data. However, conventional imaging data fails to fully represent the complex real-world scene and fails to provide sufficient information needed by deep CNNs models. These learning-based methods, therefore, demand a drastic increase in data, and more importantly, labeled data, in the case

The associate editor coordinating the review of this manuscript and approving it for publication was Wenbing Zhao.

of supervised learning to outperform the conventional lane detection methods [5]. Therefore, much of the progress in deep learning, specifically in supervised deep-network learning, can be attributed to the availability of huge image datasets such as ImageNet [6], ActivityNet [7], MS COCO [8].

Alternatively, computational imaging techniques extract richer, perhaps more perceptually meaningful information of the real-world scene with slight modification in the camera's sensors, optics, or illumination. Combining deep learning models with computational imaging enables us to exploit the full potential of image content, and therefore improve the performance of these models, particularly for environmental perception tasks.

In the field of computational imaging, light field (LF) imaging, has enabled the acquisition of angular information of the environment, in addition to spatial information in conventional imaging, by separately recording the intensities of light rays coming from different directions at each separate pixel position [9], [10]. This additional angular information can be a critical discriminatory factor for various classification/detection tasks. For example, in the case of lane detection, the light reflected from one color lane line refracts at a different angle by the lens compared to the light coming from another color lane line or the dark road from the same depth. These different angles of incidence, shown in Figure 1(c), are recorded by a light field camera which serves as an effective cue for lane detection.

Light field imaging systems can be implemented in a variety of ways including, microlens array (MLA) based cameras [10], [11], coded mask LF camera designs [12], [13], and camera arrays [14], [15]. MLA-based LF cameras are the most cost-effective option and, thereofore, adopted in this paper for data acquisition. Light fields can be used simply as perspective images or represented in different ways, such as focused stacked image, epipolar plane image (EPI), or disparity map extracted from these images.

In this paper,[1] we propose using light fields instead of the regular 2D images to train the state-of-the-art deep learning models for road lane detection aiming to improve the overall prediction accuracy with a minimal training dataset. For this purpose, we proposed two different representations of the light field, a sequence of perspective images, and a disparity map extracted from the light field as a replacement of one of the color channels of a regular image. We have trained two different network architecture designs to use these two different representations. For sequence input of perspective images we have used CNNs as a feature extractors and fed the sequence features to an LSTM network to exploit the angular relationship between these perspectives. Whereas, for regular image with the disparity map we have utilized pre-trained classification CNNs models and modified them for regression.

[1]This research is funded by the Natural Sciences and Engineering Research Council of Canada and the Canada Research Chair Program.

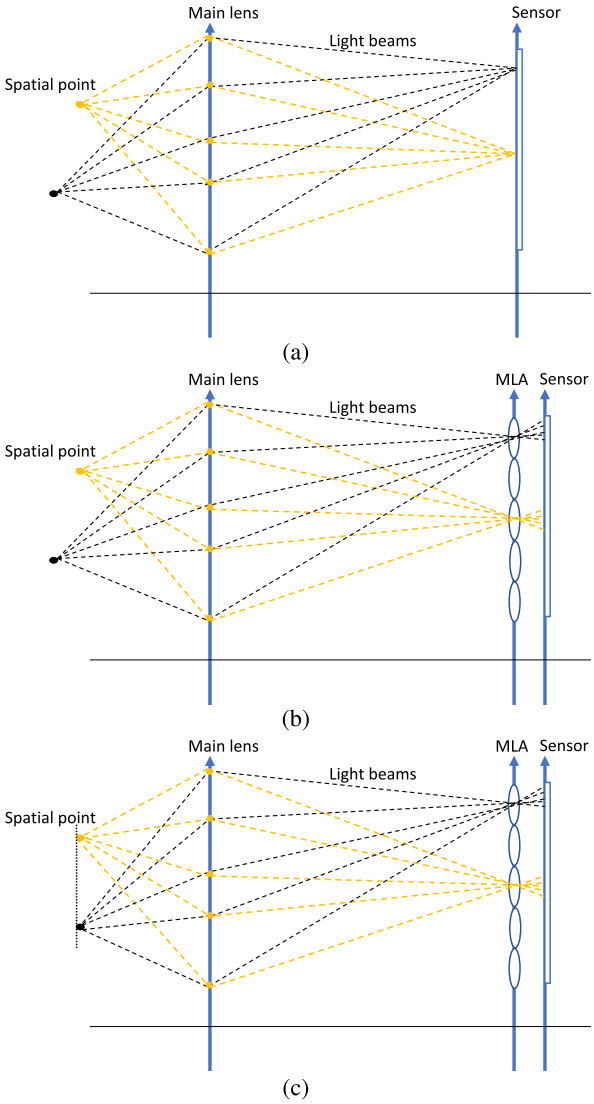

FIGURE 1. An illustration of MLA-based light field camera image acquisition in comparison with conventional camera model. (a) Conventional camera model. (b) MLA-based light field camera capturing scene points at different depths. (c) MLA-based light field camera capturing scene points at the same depth.

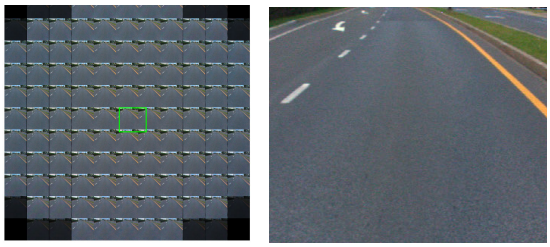

FIGURE 2. A decoded light field using Matlab's light field toolbox [16], along with a middle perspective image.

The major contributions of this paper are summarized as follows:

1) Light fields are utilized for the first time in lane detection to improve the performance of existing state-of-the-art lane detection methods with minimal training data.

2) A novel disparity cue originating only from the lane line color and independent of the depth is proposed for robust discrimination between different color lane lines and the road.
3) A road lane light field dataset is provided for further research on the use of LFs for lane detection.

The paper is organized as follows. Section II presents related work on lane detection, particularly CNN-based approaches for lane detection and recognition. In section III we present the proposed light field road lane dataset. We explain the application of light field to the lane detection problem in Section IV. Light field representations used in this paper for training deep learning models are presented in section V. In section VI we detail the training process and hyper-parameters. In sections VII and VIII we present the experimental result and provide a detailed discussion and future direction respectively. Finally, we conclude the paper in section IX.

II. RELATED WORK

Lane detection is an extensively studied topic and a vast literature exists on addressing the limitations associated with it. Different pre-processing steps such as noise reduction [17], [18], and illumination invariance [19], [20], have shown improvement in heuristic recognition-based lane detection. Feature extraction also plays a critical role in the performance of a lane detection algorithm. Some of the famous feature extractors used for lane detection are Sobel [21], Canny [22], and Hough transform [23].

The advent of convolution neural networks has revolutionized the performance of machine vision tasks and CNNs based methods have started to become the state-of-the-art. The application of CNNs specific to lane detection problem can be categorized into three types: image classification, object detection, and semantic segmentation. In classification methods, some prior information is utilized in the determination of lane positions, whereas in object detection style methods coordinate regression is employed to determine the feature points. Lane detection using semantic segmentation simply performs pixel-level classification of the lane and background pixels.

In [24] the CNN is combined with RANSAC for the first time for lane detection problem. The role of a simple eight layers CNN with three convolutional, two sub-sampling, and three fully-connected layers, is to extract features while RANSAC performed clustering. Deeplane [25] has extended depth as compared to [24], and introduced the additional normalization and dropout layers in their architecture design to achieve an overall improvement in the performance. A trend of increase in networks depth and width to improve the prediction accuracy is followed by AlexNet [26], GoogleNet [27], VGG [28], and Nasnet [29].

In [30] it is demonstrated that learning more than one task such as classification and detection simultaneously improves performance. State-of-the-art CNN-based detection methods YOLO [31], G-CNN [32], and SSD [33] also perform simultaneous learning and directly produce the class probability and coordinate value and excludes the region proposal stage altogether. EELane [34] jointly learned the lane and vehicle detection, where a vehicle class has a five-point output response, the first four represent the bounding box coordinates and the fifth point indicate it's depth, whereas for the lane class the output response is a six dimension vector. The first four dimensions are reserved for line segment endpoints and the last two represent the distance of the endpoint from the camera. In [35] lane detection is modeled as multiple row-wise classifications, where a two-stage module expresses features for classification. The first-stage layers jointly compress and model the horizontal components of all lanes, and the second-stage layers separately model each lane marker and directly output the lane marker vertices. In [36] affinity fields based clustering technique is proposed. These affinity fields (horizontal and vertical) enable unique lane instances to be identified. In [37], it is hypothesized that because lanes follow a regular pattern and are highly correlated, the global information can be critical in obtaining their position and, therefore, propose a novel anchor-based attention mechanism that aggregates global information.

In VPGNet [38] lane detection is combined with road marking recognition and the vanishing point, responsible for guiding both tasks, contributed significantly to the performance improvement. In [39] spatial and temporal dimensions are jointly exploited by a lane detection network in three steps; pre-processing CNN-based classification, regression, and lane fitting. Inspired by point cloud instance segmentation, [40] stacked hourglass network to predict key points on traffic lines where each key point is distinguished into an individual instance. CondLaneNet [41] propose row-wise formulation for optimizing lane line shape. Additionally, it utilizes the Recurrent Instance Module (RIM) to improve the detection of dense and fork lines. In [42] a Cross Layer Refinement Network (CLRNet) is proposed to utilize low- and high-level features for lane detection. Initially, detection is performed on high semantic features followed by refinement on fine-detail features to improve the precision.

Semantic segmentation is another machine vision application frequently applied for lane detection and recognition. In [43] an optical flow network used to determine the key and non-key frames is combined with a semantic segmentation network that performs key-frame segmentation. This is followed by the density-based spatial clustering of applications with noise (DBSCAN) to discriminate lanes and finally a mapping method is used to map pixel coordinates to camera coordinates. An encoder and decoder model is adopted in [44], the decoder's max-pooling layer index function is used to upsample the encoder through counter pooling approach to achieve semantic segmentation

In [45] a semantic segmentation-inspired approach is used to minimize computations in the decoder by using multiple level features of the encoder and reduced pixel embedding branches which ensures the effective utilization of these

multilevel features, resulting in the overall improved prediction precision. In [46] a joint monocular camera and LIDAR-based semantic segmentation system is proposed for road and lane marking detection. The joint system is calibrated and the LIDAR measurements are successfully used to determine the distance of the segmented road edges and lane markers from the video feed. In [47] the conventional layer-by-layer convolutions are generalized to slice-wise convolution in feature maps. This allows message exchange between pixels across rows and columns, which ensures better performance for long continuous shape structures, e.g. road lane lines.

All aforementioned learning-based methods, including the state-of-the-art utilize different CNN models as backbone. The proposed method also uses three different CNN architectures, including the lowest performing Googlenet to medium and best performing VGG-16, and Nasnet respectively, in an object detection manner and performs coordinate regression to predict the lane line points. Chromatic aberrations are used as a disparity variation between the black road and colored lane lines for the first time in lane detection. We introduced disparity as an additional cue for better road and lane line discrimination. The proposed cue has demonstrated significant performance improvement, especially in challenging conditions such as illumination variations, shadows, and false lines. Also, we have combined CNN with the LSTM to extract the angular information from the sequences of perspective images. Such an architecture has never been used to exploit disparity in lane detection. The paper lay the foundation for using LFS for lane detection problems and offer discriminative features for better predictions. Two different ways, LF color disparity, and LF LSTM are used to demonstrate the superiority of LF representations. As CNN learns to extract relevant features through training it can benefit any CNN-based lane detection approaches ranging from classification to semantic segmentation.

III. LIGHT FIELD ROAD LANE DATASET

This paper for the first time presents a light field road lane detection (LFLD) dataset. The benchmark lane detection datasets, such as CULane [47], Tusimple [48], and LLAMAS [49] consist of regular images and, therefore, only retain the spatial information of complex real-world scenes. Unlike these standard datasets, LFLD consists of light fields and hence contains the angular information of the light rays in addition to the spatial information. The additional information carries robust features for classification/detection tasks.

LFLD dataset consists of 1000 light fields captured across several roads in daylight conditions. It consists of 50 different sequences, each containing 20 light fields captured along a single road section. These sequences are densely sampled, so the dataset can be beneficial for lane detection methods that exploit temporal information (video-based lane detection method) as well as spatial and angular information. One of these sequences is shown in Figure 3, and some randomly selected captures from multiple other sequences are given in Figures 4, 5, 6, and 8. Out of the total 1000 light fields 100 are captured under challenging conditions such as, illumination variations, false lane lines and strong shadows. The remaining 900 light fields are captured under normal conditions.

LFLD dataset is captured by the first generation commercially available Lytro [50], light field camera. The captured light fields are decoded to obtain a set of 11×11 perspective images, each with the size 375×375 pixels, using Matlab's Light Field Toolbox v0.4 [16], as shown in Figure 2. The dataset has been made public and is available on the link given below.[2]

IV. LIGHT FIELD FOR IMPROVED LANE DETECTION

Light field imaging captures the angular information of the light rays originating from a scene point. The MLA based light field cameras, Figure 1(b & c), achieve this by placing an MLA adjacent to the sensor that separately records at different pixels the intensities of light rays passing through different points (sub-apertures) of the main lens and converging at the micro-lens in front of these pixels. However, in conventional cameras, Figure 1(a), the incident light rays converge directly at the sensor and are recorded, therefore, losing the directional information.

Typically, the light rays reflected from scene points at different depths arrive at the MLA with a different angle, Figure 1(b). By tracing light rays back to the scene space, the corresponding depth of these points can be calculated. Depth estimation of all the captured points results in a map that provides the depth variations among different objects in the scene. This depth information is an additional discriminative feature well utilized in classification problems such as face recognition.

In case of the road lane detection, depth of the lane lines on the road changes along with the road's depth. Therefore, the scene's depth variation does not provide any valuable information to improve the discrimination between different lane lines and the road. However, color of the lane lines is unique and significantly differs from the road's color as well. Light reflected from the unique color (wavelength) lane line, and the road from the two different points at the same depth passes from the main lens and refracts following the refraction theorem. Therefore, the angle of incidence at MLA of light rays originating from different color points despite the same depth is still different, as shown in Figure 1(c). This variation in incident angle cause disparity in multiple perspective images acquired from the decoded light field.

In Figure 4, it can be seen that the disparity changes with the change in color of the lane line. This disparity variation provides an additional discrimination cue that can improve the performance of lane detection methods. It is also worth noticing that despite the change in depth along the road, the disparity seems to be mostly constant. This could be due to the fact that the black road is uniform and there is not enough texture variation on the road for any matching algorithm to

[2]https://github.com/LotfiZ/LFW_database

FIGURE 3. Example sequence of the proposed LFLD dataset.

FIGURE 4. Sample light field middle perspective images from our road lane dataset and their corresponding disparity maps. (Row 1) Sample road lane perspective images. (Row 2) Corresponding disparities maps.

detect. This significantly improves the effectiveness of the proposed disparity cue, ensuring that disparity variation is only caused by the presence of lane lines on the road.

Shades of the road trees and vehicles on the road usually result in occlusion and illumination variation across the lane lines. False lines can also appear on the road, for example, shades of high tension wires along the road, as shown in Figure 5. This causes the intensity variation along the lane lines in the captured image and therefore limits their effectiveness in providing sufficient discriminative features for detection. However, the wavelength of the light reflected from a particular color line remains independent of the illumination variation and therefore the angular information remains unaffected.

An EPI image can be extracted from a light field by cross-sectioning correspondingly orientated perspectives. For instance, gathering pixels by horizontally cross-sectioning perspectives in horizontal direction results in horizontal EPI. Similarly, vertical EPI can be formed through vertical cross-sectioning of the vertically aligned perspectives. The EPI representation of the light field, Figure 6, shows that the angle of the incident light across a lane line remains

FIGURE 5. A subset of various challenging conditions for lane detection from our challenging road condition dataset.

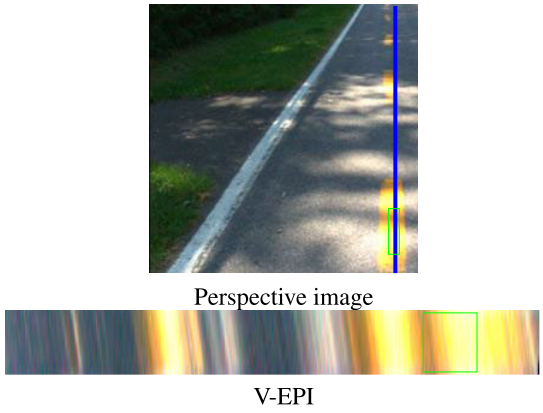

FIGURE 6. An EPI-based illustration of light's incident angle under illumination variations.

unchanged irrespective of the illumination variation. The additional angular information provides robustness against illumination variations and false lines. Multiple viewpoints help significantly in occlusion avoidance, and therefore, the degradation in the overall performance of any lane detection algorithm under these challenging conditions is minimal with the light field as input data.

V. LIGHT FIELD REPRESENTATION

Light field can be represented in many ways including, perspective images, lenslet images, EPIs, and focus stacked images. In this paper, we used light fields in two different forms, a sequence of perspective images and a disparity map estimated through the light field, to take advantage of additional angular information for improved lane detection. Other light field representations, such as Lenslet, EPI, and focused stacked image, can be considered by the future lane detection methods that will incorporate LFs as their input.

A. PERSPECTIVE IMAGE REPRESENTATION

Lytro camera's raw light field when decoded with the [16], Matlab's toolbox results in a regular grid of 11 × 11 perspective images which is referred to as its angular resolution and each perspective has a spatial resolution of 375 × 375. We have cropped and resized these perspective images to match CNN's input layer size. In order to exploit the angular information among the perspective images, we used an LSTM layer in our network shown in Figure 7, and detailed below.

An LSTM layer is composed of cells and each cell has three inputs, an input feature vector, an input hidden state, and a common cell state. Additionally, each cell has three gates: input, forget, and output gates. Upon the arrival of new information, the network can forget the previous state and update the current state. This structure enables learning long short-term inter-view, angular relationships from the features extracted through the sequence of perspective images.

The overall network architecture includes a convolutional neural network, which is used as a feature extractor and converts a sequence of perspective images into a sequence of feature vectors, where each feature vector is obtained from the last pooling layer of the network independently for each input perspective image and then combined as a sequence. The sequence of feature vectors is fed directly to a sequence input layer with an input size corresponding to the feature dimension of the feature vectors. An input sequence layer is followed by an LSTM layer with 2000 hidden units and a dropout layer afterwards. Finally, a fully connected layer with an output size matching the number of responses and a regression layer are connected.

B. DISPARITY MAP AS INPUT

The variation in the angle of incident light originating from different color lane lines and roads can be represented in the form of a disparity map. Therefore, it is possible to utilize this angular information by directly providing a disparity map as input to the CNN. Since the lane lines are the only color variation on the road data, we assume that using the correct RGB primaries would allow us to represent the color information adequately. For example, only red and green channels can represent a yellow line. However, only in case of the white line, all three color channels are needed and, therefore, the information from one color channel for one lane line is traded for the additional information from disparity for all road lines.

In our experiment, we used the Lytro desktop app[3]to estimate disparity maps. The Lytro app allows the export of gray-scale 16-bit normalized disparity maps from a given LF image. Smoothing is applied to the disparity map by the software with the help of human interaction for noise reduction. Since the entire LF is used in estimating the disparity maps, it is expected to be robust to the matching errors in comparison to the typical stereo systems. We have tested two combinations: discarding a single color channel (blue) of an RGB image and replacing it with the disparity map, and converting an RGB image into gray-scale and replicating the same information in two channels and using the disparity map as the third channel, without changing the input layer to

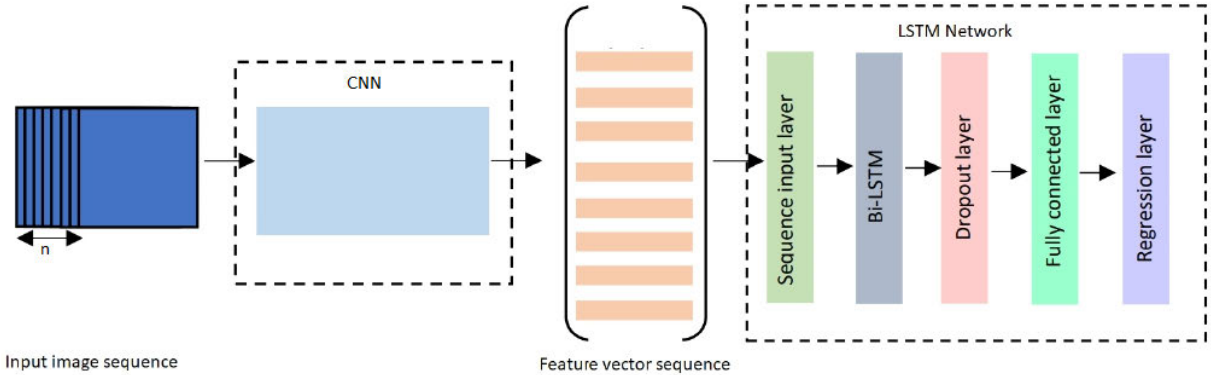

FIGURE 7. An illustration of network architecture designed for utilizing the inter-view angular information of the light field's perspective image representation.

use pre-trained weights. Although this representation results in the information trade-off, it allows us to avoid the extra computation and network modification involved in the LSTM based approach. However, this approach involves an additional pre-processing step of estimating the disparity map. The proposed representation can be used with any existing network for lane detection.

VI. TRAINING

We trained three different CNNs, namely GoogleNet, VGG-16, and Nasnet. To demonstrate the robustness of the proposed LF representation, we chose these networks which reflect a wide range of performance variation in terms of prediction accuracy on ImageNet [6] validation dataset.

These CNNs are classification networks and for the road lane line points coordinate prediction, we modified them for regression. For GoogleNet, we replaced the 'loss3-classifier', 'prob', 'output' layers with a fully connected (FC) layer with 20 responses, to match the coordinate of the lane line points per image and in the end, we added a regression layer. For VGG-16 we removed the fully connected, 'prob', 'output' layers and added an FC and regression layer. For Nasnet, we removed the final prediction layers and the classification layer and introduced an additional FC and regression layer.

These CNNs are pre-trained on ImageNet dataset and transfer learning is performed to adapt and fine-tune pre-trained GoogleNet, VGG-16, and Nasnet for the lane detection problem. For the sequence of perspective images, Pre-trained CNN is used for feature extraction on our road data and LSTM network pre-trained on hmdb [51], data is fined tuned for the lane detection problem. Our road lane dataset, which is captured across different cities, consists of 1000 light fields, out of which 70 % are used for training and the remaining 30% is kept for testing.

The simulations show that the network's predictions are robust under different training hyper-parameters configurations. Various combinations of the learning rate, optimizer, and batch size are tested but the variations in performance are negligible. The batch size turned out to be the most influential hyper-parameter in terms of performance improvement for both the networks described above. Increasing the batch size improved the performance and the maximum batch size supported due to memory limitations is 32.

Among different configurations, the following set of hyperparameters resulted in the best performance and, therefore, adopted in this work. The initial learning rate is set to 3e-4 with a learning rate drop of 0.1 and a learning rate drop period of 20. Although the choice of optimizer has shown a marginal effect on the overall performance, adam optimizer has slightly improved the prediction and hence selected in our training. The batch size is set to 32 and the training is performed for 40 epochs.

A. EVALUATION METRICS

1) RMSE

Evaluation of the regression model's performance can best be described in terms of error values. For this purpose, we calculated the commonly used root mean square error metric (RMSE) [52] and [53]. The following equation presents the formula to calculate RMSE.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Actual_i - Predicted_i)^2}{N_{images}}}, \tag{1}$$

where $Predicted_i$ represents the predicted lane line point coordinates value in pixels units, and $Actual_i$ is the ground truth value for the same point coordinates. In practice, Actual and Predicted values are given as two-dimensional vectors with x and y coordinates of the selected points on the lane lines. So the smaller RMSE will indicate that the predicted lane line is closer to the ground truth than a predicted line with a large RMSE.

2) LIOU

We have also used Line IoU loss [42], where possible, for the evaluation of the proposed disparity-based LF representation. Each point x_i^p in the predicted lane is first extended into a line segment with a radius e. Then IoU, which is the ratio of intersection over union between two line segments, is calculated between the extended line segment and its ground truth,

[3]http://lightfield-forum.com/lytro/lytro-archive/

FIGURE 8. Visualization of the predicted lane lines on a subset of dataset. (Row 1 & 2) Examples of accurately predicted lanes. Row 3) Examples of poorly predicted lane lines.

as given below:

$$IoU = \frac{d_i^{\mathcal{O}}}{d_i^{\mathcal{U}}} = \frac{\min(x_i^p + e, x_i^g + e) - \max(x_i^p - e, x_i^g - e)}{\max(x_i^p + e, x_i^g + e) - \min(x_i^p - e, x_i^g - e)}, \tag{2}$$

where $x_i^p - e, x_i^p + e$ are the extended points of x_i^p, $x_i^q - e, x_i^q + e$ are the corresponding ground truth points. Note that $d_i^{\mathcal{O}}$ can be negative, which can make it feasible to optimize in case of non-overlapping line segments. Then LIoU can be considered as the combination of infinite line points. To simplify the expression and make it easy to compute, we transform it into a discrete form,

$$LIoU = \frac{\sum_{i=1}^{N} d_i^{\mathcal{O}}}{\sum_{i=1}^{N} d_i^{\mathcal{U}}}. \tag{3}$$

Then, the LIoU loss is defined as:

$$\mathcal{L}_{LIoU} = 1 - LIoU \tag{4}$$

where $-1 \leq LIoU \leq 1$, when two lines overlay perfectly, then LIoU = 1, LIoU converges to -1 when two lines are far away.

3) ACCURACY

Accuracy is another commonly adopted metric for the evaluation of a lane detection model [54], [55]. We applied accuracy for the performance evaluation of the proposed disparity-based LF representation using state-of-the-art lane detection methods. The formula for accuracy estimation is given below.

$$Accuracy = \frac{\sum_{clip} C_{clip}}{C_{clip} S_{clip}} \tag{5}$$

where C_{clip}, S_{clip} are the number of correctly predicted lane line points and the number of ground truth lane line points, respectively. A predicted lane is considered correct if more than 85 % of predicted lane line points are within 20 pixels of the corresponding ground truth points.

VII. EXPERIMENTAL RESULTS

This section presents the results of the proposed light field-based lane detection technique. We have evaluated the proposed technique quantitatively and qualitatively by using the evaluation metrics detailed in subsection VI-A, and plotting the predicted lane lines compared to the ground truth for visual inspection. We chose only the conventional RGB image to plot for simplification of representation; the lane lines predicted using different light field representations and 2D images against the ground truth.

In Figure 8, it can be seen that in both good and poor quality predictions, the LF-based lane detections are overall closer to the ground truth as compared to the regular image-based predictions. However, in this small subset presented in Figure 8, it is unclear that a particular LF representation performs better than the others.

In Figure 9, a quantitative comparison of the different LF representations and regular images (middle perspective image of the decoded LF) is presented. It can be seen in Figure 9, that any LF representation outperforms the regular image-based lane detection by a significant margin. Within different LF representations, a sequence of input perspective images has shown slight improvement over the others. Nevertheless, the sequence-based input requires modification in the network architecture and involves additional computations; however, no additional labeling is required.

The LF perspective image representation, however, does not benefit solely from the increased number of images. The most critical factor is the angular information (Disparity) between these perspectives. In Figure 10, it can be seen that once the angular information between the perspective

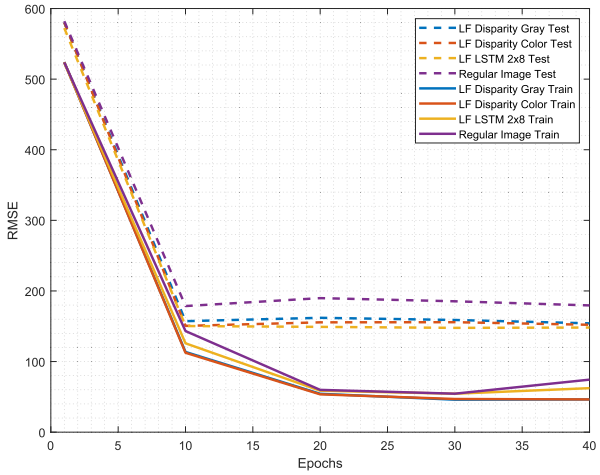

FIGURE 9. Performance comparison of different LF representations and conventional images on testing and training datasets using googlenet.

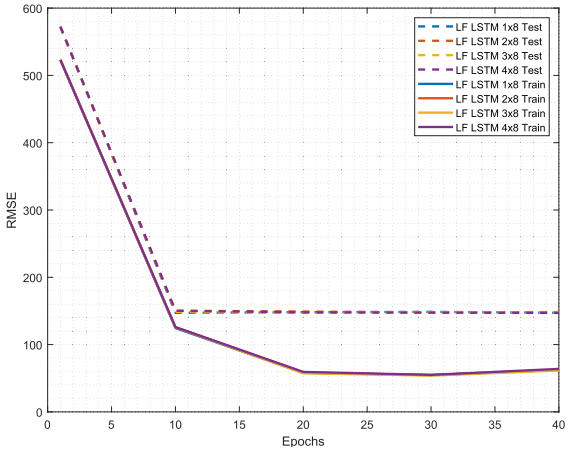

FIGURE 10. Performance analysis by increasing the number of perspective images in the input sequence of the LSTM network.

images is learned by the network, increasing the number of views makes negligible improvement in the performance. We have noticed that up to eight perspective images are sufficient to adequately present the angular information in our case. The impact of viewpoint selection is not considered in this study and will be part of future work along with other possible compact representations. Currently, the horizontal perspective images from the middle row of the decoded light field, Figure 2, are selected. To increase the views points, as presented in Figure 10, the rows on top and underneath the middle row of the decoded light field are added.

The robustness of the proposed light field-based lane detection method in the challenging conditions is evident from Figure 11. In our experiments, the following conditions: illumination variation, shadow, and false lines, highlighted in Figure 5, constitute the challenging dataset. Under these special conditions, the spatial information gets degraded, resulting in performance deterioration of the methods that rely only on the spatial information of the scene. The LF representations provide additional angular information independent of the illumination variations. Multiple views are

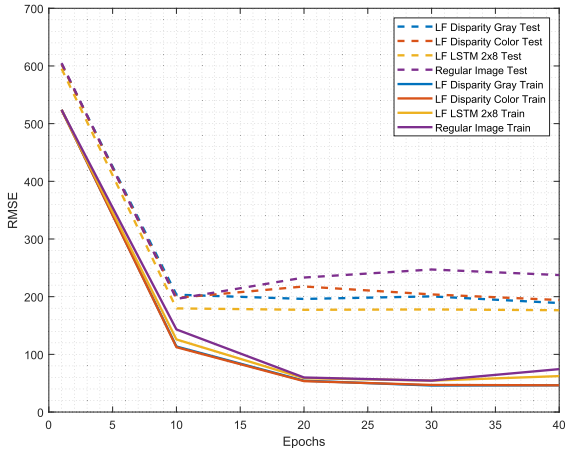

FIGURE 11. Comparison of proposed LF representations and regular images on a challenging test dataset.

effective in the case of occlusions, and the disparity provides better discrimination against false lines. For the results presented in Figure 11, the network is trained on a regular condition training dataset of 700 images, and only testing is performed on the challenging dataset of 100 samples. It can be seen in Figure 11, that the challenging conditions dataset despite being $\frac{1}{3}$ in size, results in overall higher RMSE as compared to the testing performed on a regular condition dataset presented in Figure 9.

In Table 1, we compare the performance of the LF representations and regular images on two different test datasets. The proposed LF representation (LF Disparity Gray, LF Disparity Color, LF LSTM 2×8) outperforms the regular image data by 14%, 15%, and 18% under normal conditions; however, the difference in performance grows to 20%, 18%, and 26% in challenging conditions. The difference in performance is expected to increase three times given two different conditions have the same number of test samples.

In Figure 12, we demonstrate that a CNN model designed for lane detection can benefit primarily from using LFs during training. It can be seen that a network trained on regular images needs additional 200 images to achieve comparable performance to the network trained using LFs. The difference could be more significant if the network is trained from scratch instead of fine-tuning a pre-trained network. This performance improvement with lesser data can benefit scarce data-set and avoid additional labeling costs.

Several deep CNN architectures designed for image classification exist in the literature. Many design choices influence the performance of these CNN models, for example, the number of trainable parameters, Depth, and width of architecture [56]. To demonstrate the robustness of the proposed method across different CNN architectures, which constitute a significant part of most state-of-the-art lane detection methods and is used as a feature extractor in novel learning-based methods. We have compared the performance of LF representation with the regular images on two other CNN architecture, namely the VGG-16, and Nasnet presented in Figure 13 and 14.

TABLE 1. Quantitative performance comparison of the proposed LF representation and the regular image approach for lane detection.

Input data	Test rmse					
	Epochs					
	1	10	20	30	40	Avg.
LF Disparity Gray Test	580.1897	157.1864	161.9565	158.8658	154.1153	242.4627
LF Disparity Color Test	580.1895	150.4955	155.5348	155.7526	152.0849	238.8115
LF LSTM 2x8 Test	572.3088	150.3050	149.1151	147.7085	148.1758	233.5226
Regular Image Test	582.0514	178.5948	189.8213	185.3534	179.5123	263.0666
Challenging conditions						
LF Disparity Gray Test	603.0599	203.4266	196.0821	200.5210	188.9750	278.4129
LF Disparity Color Test	603.0569	197.9583	217.7106	203.9217	194.1931	283.3681
LF LSTM 2x8 Test	595.0994	179.6736	177.2515	177.8729	176.5384	261.2872
Regular Image Test	604.9989	196.1004	233.1878	247.1022	237.4779	303.7734

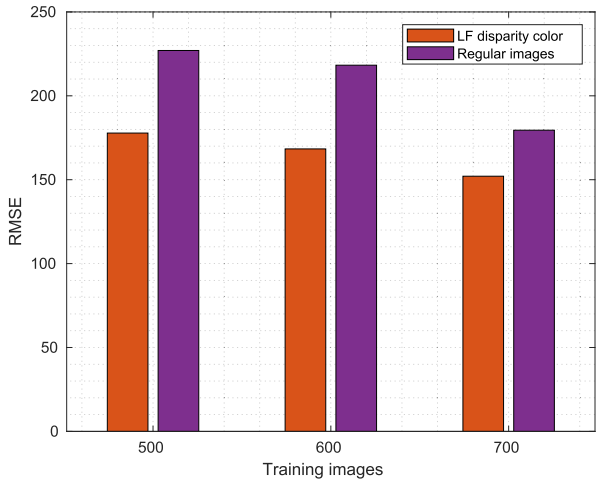

FIGURE 12. Performance comparison on reduced number of training images. Network trained on least number of LFs has comparable performance to the network trained on maximum number of regular images.

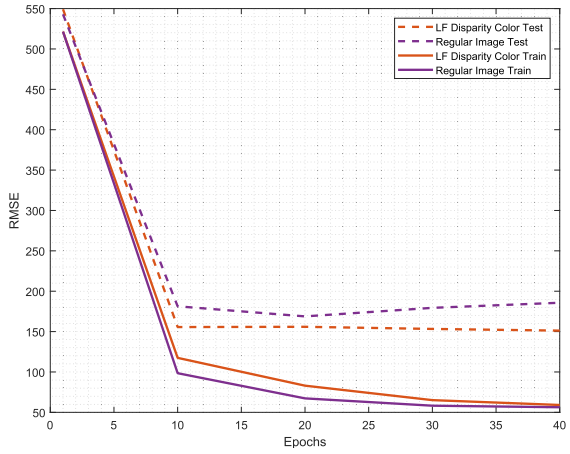

FIGURE 13. Robust evaluation of the proposed LF approach and performance comparisons with traditional image-based technique on a different (VGG-16) CNN architecture.

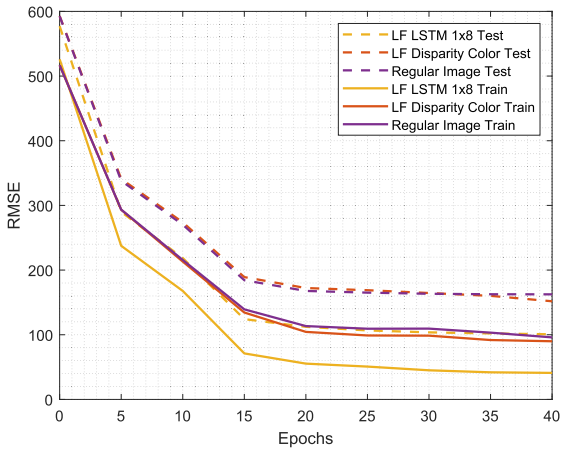

FIGURE 14. Robust evaluation of the proposed LF approach and performance comparisons with traditional image-based technique on another (Nasnet) CNN architecture.

The performance gap between the VGG-16 and GoogleNet is marginal; Nasnet, in combination with LSTM trained on the LF data, achieves significant performance improvement. However, for Nasnet, the gap between regular image and disparity-based LF representation reduces only to 6.3 % as Nasnet shows improvement in performance over GoogleNet and VGG-16 for the regular image dataset. As we used the transfer learning technique to fine-tune the pre-trained models, the networks are biased towards the regular image dataset due to the limited size of LF training data. Since LSTM allows us to maintain the regular RGB image representation and use multiple perspective images to exploit the angular information, it avoids the network's bias. Increasing the data set size for fine-tuning could further increase the performance gap between the disparity-based LF representation and regular image dataset even for complex networks like Nasnet.

Although the three networks are widely different in their design approach, it can be seen that the proposed LF representation in general out perform the regular image based approach. Therefore, the proposed light field-based road lane representation can benefit any existing lane detection method and possibly a new lane detection method such as combining CNN with LSTM or any other network architecture that can exploit the angular information.

In Table 2, we compared the performance of the proposed light field disparity representation with regular images on state-of-the-art methods. The proposed LF representation performs better than the traditional images on lane

TABLE 2. Comparison of the proposed LF disparity representation with regular images on state of the art lane detection methods.

Method	Backbone	Pre-trained ImageNet [6]	Fine-tuning	Accuracy	LIOU loss
CLRNet [42] [2022]	ResNet18	✓	LFLD (regular images)	54.3%	0.249
CLRNet [42] [2022]	ResNet18	✓	LFLD (LF disparity)	56.6%	0.194
LaneATT [37] [2021]	ResNet18	✓	LFLD (regular images)	44.7	–
Lane ATT [37] [2021]	ResNet18	✓	LFLD (LF disparity)	46.5	–

detection methods explicitly designed for convectional images. It should also be noted that the backbone networks are pre-trained on Imagenet [6], which consists of millions of regular images and, therefore, are biased towards conventional images. The performance gap is expected to increase further with the addition of more training samples in the LFLD dataset during the fine-tuning of the state-of-the-art methods. Additionally, it can be seen in Table 1 and Figure 14, that LF perspective image representation, when used with LSTM, supersedes the LF disparity representation by a significant margin. Therefore, it can be concluded that LF representations not only benefit state-of-the-art lane detection methods but are also worth investigating new methods explicitly designed for exploiting the angular information in addition to spatial information of the scene.

VIII. DISCUSSION AND FUTURE WORK

This paper proposes a spatio-angular learning framework for light field lane detection. In the perspective image-based LF representation approach, an LSTM network, that takes as input a sequence of GoogleNet CNN features extracted for each perspective independently, is used to learn the inter-view, angular information present in multiple perspective images of a light field. It can be seen in Figure 10, that increasing the perspective images in the input sequence above eight viewpoints has minimal effect on lane line coordinate prediction. However, it should be noted that the perspective images are selected along the horizontal axis (excluding the corner viewpoints to avoid vignetting effect) of the decoded light field. Along the vertical axis, only up to 4 adjacent viewpoints to the middle perspectives are utilized in input sequences with more than eight viewpoints case to limit the excessive additional computational cost. In the future, we plan to investigate the effective selection of horizontal and vertical perspectives to further exploit the additional information about the inter-view angular relation, without significantly increasing the computational cost.

LF is utilized in a simple pre-processing step for disparity (the proposed cue for improved lane detection) estimation to avoid any modifications and extensions of the existing CNN-based lane detection networks. A trade-off between the color information and the proposed angular information exists in this approach. Since the road data is not color intensive and the only color variations are lane lines, careful selection of the RGB color primaries can represent most of the scene's color information. In Figure 9, it can be seen that the performance gap between disparity map-based and sequence of perspective image-based LF representation is marginal. However, in adverse conditions, Figure 11, the difference in the performance of the two representations is obvious.

Light field can be represented in several ways, including EPI, focus-stacked image, set of perspective image, and lenslet image. In these different LF representations, the angular information can be extracted by either existing lane detection methods or new methods that could be developed specifically for a particular representation. We would like to investigate lenslet and EPI LF representations and their application to lane detection problems in future work.

Another direction for the extension of this work that we will consider in the future is adopting an appropriate lens for the light field imaging system, which can maximize the refractive index variation for the increased disparity between different color lane lines. The refractive index of typical transparent materials, such as the glass in lenses, is inversely related to the wavelength passing through it. Glass lenses bend different color light rays at different angles. For example, blue rays bend more than red rays. With a simple lens, red light focuses behind green light, and blue light focuses in front of green light. This phenomenon is called chromatic aberration (CA), which results in poor color registration.

Lens manufacture usually accounts for the chromatic aberrations in many lenses to achieve a single point of focus for each wavelength. Chromatic aberrations still occur for faster lenses, especially when capturing high contrast areas such as our road dataset with color lines on dark background. In the future, we would like to use a high depression lens or some low-quality lens that does not account for CAs, to increase further the disparity between lane lines and improve prediction accuracy.

IX. CONCLUSION

The paper presents a LF-based lane detection method for improved prediction accuracy. The proposed method leverage the additional angular information recorded by the light fields, specifically in challenging conditions such as illumination variation, shadows, false lane lines, and occlusion. Both LF representations presented in this paper show significant performance improvement compared to the traditional image-based lane detection approach. Although LF perspective image representation has some advantage over the other LF representation in terms of RMSE, it incurs additional computational cost and involves network design changes. The proposed approach improves lane detection by providing additional features that CNNs extract through learning. It can enhance any lane detection method that includes CNN

as a feature extractor as well as the non-learning-based approaches with careful consideration towards the hand-crafted features.

ACKNOWLEDGMENT

This research is funded by the Natural Sciences and Engineering Research Council of Canada and the Canada Research Chair Program.

REFERENCES

[1] E. Bellis and J. Page, "National motor vehicle crash causation survey (NMVCCS) SAS analytical users manual," Nat. Highway Traffic Saf. Admin., USA, Tech. Rep. DOT HS 811 053, 2008.

[2] J. E. Gayko, "Lane departure and lane keeping," *Handbook Intell. Vehicles*, vol. 1, pp. 689–708, Jan. 2012.

[3] C. Visvikis, T. Smith, M. Pitcher, and R. Smith, *Study on Lane Departure Warning and Lane Change Assistant Systems*, vol. 374. Crowthorne, U.K.: Transport Research Laboratory Project Rpt PPR, 2008.

[4] D. Hopkins and T. Schwanen, "Talking about automated vehicles: What do levels of automation do?" *Technol. Soc.*, vol. 64, Feb. 2021, Art. no. 101488.

[5] J. Kim and C. Park, "End-to-end ego lane estimation based on sequential transfer learning for self-driving cars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 30–38.

[6] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[7] F. C. Heilbron, V. Escorcia, B. Ghanem, and J. C. Niebles, "ActivityNet: A large-scale video benchmark for human activity understanding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 961–970.

[8] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 740–755.

[9] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. 23rd Annu. Conf. Comput. Graph. Interact. Techn. (SIGGRAPH)*, 1996, pp. 31–42.

[10] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Stanford Univ. Comput. Sci., Stanford, CA, USA, Tech. Rep. CSTR, 2005.

[11] M. Z. Alam and B. K. Gunturk, "Hybrid light field imaging for improved spatial resolution and depth range," *Mach. Vis. Appl.*, vol. 29, no. 1, pp. 11–22, 2018.

[12] A. Ashok and M. A. Neifeld, "Compressive light field imaging," *Proc. SPIE*, vol. 7690, May 2010, Art. no. 76900Q.

[13] M. Z. Alam and B. K. Gunturk, "Deconvolution based light field extraction from a single image capture," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 420–424.

[14] B. Wilburn, N. Joshi, V. Vaish, E. V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 765–776, Jul. 2005.

[15] J. C. Yang, M. Everett, C. Buehler, and L. McMillan, "A real-time distributed light field camera," in *Proc. Eurographics Workshop Rendering*, 2002, pp. 77–86.

[16] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for Lenselet-based plenoptic cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1027–1034.

[17] J. G. Wang, C.-J. Lin, and S.-M. Chen, "Applying fuzzy method to vision-based lane detection and departure warning system," *Expert Syst. Appl.*, vol. 37, pp. 113–126, Jan. 2010.

[18] P.-Y. Hsiao, C.-W. Yeh, S.-S. Huang, and L.-C. Fu, "A portable vision-based real-time lane departure warning system: Day and night," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 2089–2094, May 2009.

[19] Y. Chai, S. J. Wei, and X. C. Li, "The multi-scale Hough transform lane detection method based on the algorithm of Otsu and Canny," *Adv. Mater. Res.*, vol. 1042, pp. 126–130, Oct. 2014.

[20] V. Gaikwad and S. Lokhande, "Lane departure identification for advanced driver assistance," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 910–918, Apr. 2015.

[21] C. Mu and X. Ma, "Lane detection based on object segmentation and piecewise fitting," *TELKOMNIKA Indonesian J. Elect. Eng.*, vol. 12, no. 5, pp. 3491–3500, May 2014.

[22] P.-C. Wu, C.-Y. Chang, and C. H. Lin, "Lane-mark extraction for automobiles under complex conditions," *Pattern Recognit.*, vol. 47, no. 8, pp. 2756–2767, Aug. 2014.

[23] J. Niu, J. Lu, M. Xu, P. Lv, and X. Zhao, "Robust lane detection using two-stage feature extraction with curve fitting," *Pattern Recognit.*, vol. 59, pp. 225–233, Nov. 2016.

[24] J. Kim and M. Lee, "Robust lane detection based on convolutional neural network and random sample consensus," in *Proc. Int. Conf. Neural Inf. Process.* Cham, Switzerland: Springer, 2014, pp. 454–461.

[25] A. Gurghian, T. Koduri, S. V. Bailur, K. J. Carey, and V. N. Murali, "DeepLanes: End-to-end lane position estimation using deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2016, pp. 38–45.

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 25. Stateline, NV, USA, Dec. 2012, pp. 1097–1105.

[27] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[28] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[29] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8697–8710.

[30] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "OverFeat: Integrated recognition, localization and detection using convolutional networks," 2013, *arXiv:1312.6229*.

[31] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[32] M. Najibi, M. Rastegari, and L. S. Davis, "G-CNN: An iterative grid based object detector," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2369–2377.

[33] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 21–37.

[34] B. Huval, T. Wang, S. Tandon, J. Kiske, W. Song, J. Pazhayampallil, M. Andriluka, P. Rajpurkar, T. Migimatsu, R. Cheng-Yue, F. Mujica, A. Coates, and A. Y. Ng, "An empirical evaluation of deep learning on highway driving," 2015, *arXiv:1504.01716*.

[35] S. Yoo, H. S. Lee, H. Myeong, S. Yun, H. Park, J. Cho, and D. H. Kim, "End-to-end lane marker detection via row-wise classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 1006–1007.

[36] H. Abualsaud, S. Liu, D. B. Lu, K. Situ, A. Rangesh, and M. M. Trivedi, "LaneAF: Robust multi-lane detection with affinity fields," *IEEE Robot. Autom. Lett.*, vol. 6, no. 4, pp. 7477–7484, Oct. 2021.

[37] L. Tabelini, R. Berriel, T. M. Paixao, C. Badue, A. F. D. Souza, and T. Oliveira-Santos, "Keep your eyes on the lane: Real-time attention-guided lane detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 294–302.

[38] S. Lee, J. Kim, J. S. Yoon, S. Shin, O. Bailo, N. Kim, T.-H. Lee, H. S. Hong, S.-H. Han, and I. S. Kweon, "VPGNet: Vanishing point guided network for lane and road marking detection and recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1947–1955.

[39] Y. Huang, S. Chen, Y. Chen, Z. Jian, and N. Zheng, "Spatial–temproal based lane detection using deep learning," in *Proc. IFIP Int. Conf. Artif. Intell. Appl. Innov.* Cham, Switzerland: Springer, 2018, pp. 143–154.

[40] Y. Ko, Y. Lee, S. Azam, F. Munir, M. Jeon, and W. Pedrycz, "Key points estimation and point instance segmentation approach for lane detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 8949–8958, Jul. 2022.

[41] L. Liu, X. Chen, S. Zhu, and P. Tan, "CondLaneNet: A top-to-down lane detection framework based on conditional convolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 3773–3782.

[42] T. Zheng, Y. Huang, Y. Liu, W. Tang, Z. Yang, D. Cai, and X. He, "CLRNet: Cross layer refinement network for lane detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 898–907.

[43] S. Lu, Z. Luo, F. Gao, M. Liu, K. Chang, and C. Piao, "A fast and robust lane detection method based on semantic segmentation and optical flow estimation," *Sensors*, vol. 21, no. 2, p. 400, Jan. 2021.
[44] L. Ding, H. Zhang, J. Xiao, C. Shu, and S. Lu, "A lane detection method based on semantic segmentation," *Comput. Model. Eng. Sci.*, vol. 122, no. 3, pp. 1039–1053, 2020.
[45] W.-J. Yang, Y.-T. Cheng, and P.-C. Chung, "Improved lane detection with multilevel features in branch convolutional neural networks," *IEEE Access*, vol. 7, pp. 173148–173156, 2019.
[46] K. L. Lim, T. Drage, and T. Braunl, "Implementation of semantic segmentation for road and lane detection on an autonomous ground vehicle with LIDAR," in *Proc. IEEE Int. Conf. Multisensor Fusion Integr. Intell. Syst. (MFI)*, Nov. 2017, pp. 429–434.
[47] X. Pan, J. Shi, P. Luo, X. Wang, and X. Tang, "Spatial as deep: Spatial CNN for traffic scene understanding," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, 2018, pp. 7276–7283.
[48] *Tusimple. Tusimple Benchmark*. Accessed: Apr. 2021. [Online]. Available: https://github.com/TuSimple/tusimple-benchmark/tree/master/doc/
[49] K. Behrendt and R. Soussan, "Unsupervised labeled lane markers using maps," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 832–839.
[50] *Lytro, Inc*. Accessed: May 24, 2018. [Online]. Available: https://support.lytro.com/hc/en-us/
[51] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, "HMDB: A large video database for human motion recognition," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2556–2563.
[52] H. U. Khan, A. R. Ali, A. Hassan, A. Ali, W. Kazmi, and A. Zaheer, "Lane detection using lane boundary marker network with road geometry constraints," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 1834–1843.
[53] R. Tapia-Espinoza and M. Torres-Torriti, "Robust lane sensing and departure warning under shadows and occlusions," *Sensors*, vol. 13, no. 3, pp. 3270–3298, Mar. 2013.
[54] Q. Zou, H. Jiang, Q. Dai, and Y. Yue, "Robust lane detection from continuous driving scenes using deep neural networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 41–54, Mar. 2019.
[55] J. Tang, S. Li, and P. Liu, "A review of lane detection methods based on deep learning," *Pattern Recognit.*, vol. 111, Mar. 2021, Art. no. 107623.
[56] M. Z. Alam, H. F. Ates, T. Baykas, and B. K. Gunturk, "Analysis of deep learning based path loss prediction from satellite images," in *Proc. 29th Signal Process. Commun. Appl. Conf. (SIU)*, Jun. 2021, pp. 1–4.

MUHAMAD ZESHAN ALAM received the B.S. degree in computer engineering from COMSATS University, Pakistan, the M.S. degree in electrical and electronics engineering from the University of Bradford, U.K., and the Ph.D. degree in electrical engineering and cyber-systems from Istanbul Medipol University, Turkey. He worked at the University of Cambridge, as a Postdoctoral Fellow, where his work focused on computer vision and machine learning models. He currently joined Brandon University, Canada, as an Assistant Professor, while also working as a Computer Vision Consultant at Vimmerse Inc. His research interests include immersive videos, computational imaging, computer vision, and machine learning modeling.

SOUSSO KELOUWANI received the B.S. and M.Sc.A. degrees in electrical engineering from the University of Quebec at Trois Rivieres (UQTR), Trois Rivières, QC, Canada, in 2000 and 2002, respectively, and the Ph.D. degree in electrical engineering (automation and systems) from the École Polytechnique de Montréal, Montréal, QC, in 2010. Before starting his Ph.D. studies, he worked in research and development in the field of cell phone application optimization at Cilys 53 Inc., from 2002 to 2005, and Openwave Inc., from 2005 to 2006. He was a holder of the Industrial Research Chair DIVEL in intelligent navigation of autonomous industrial vehicles. He has been a Full Professor of mechatronics at the Department of Mechanical Engineering, UQTR, since 2018. He was a holder of three patents in the U.S. His research interests include optimization of energy systems for vehicular applications, advanced driving assistance techniques, eco-energy navigation of autonomous vehicles, and hybridization of energy sources for vehicles with low ecological impact (battery, fuel cell, and hydrogen generator) in harsh weather conditions. In 2017, he received the Environment Award of the Grand Prix for Excellence in Transport from the Quebec Transportation Association (AQTr) for the development of a hydrogen range extender based on a hydrogen generator for electric vehicles. He was also a recipient of the Canada Governor General's Gold Medal in 2000. Moreover, he has worked with several Canadian transportation companies to develop intelligent, energy efficient, and driverless vehicles.

JONATHAN BOISCLAIR received the B.Sc.A. degree in computer science in April 2017, the M.Sc. degree from the University of Quebec at Trois Rivieres (UQTR), Trois-Rivières, QC, Canada, in 2019, where he is currently pursuing the Ph.D. degree in mechanical engineering to improve his knowledge of applied artificial intelligence. His main research interests include artificial intelligence, autonomous driving, advanced driving techniques, and real computer intelligence.

ALI AKREM AMAMOU received the B.S. degree in industrial computing and automatic science from the National Institute of Applied Sciences and Technology, Tunis, Tunisia, in 2013, the M.S. degree in embedded systems science from Arts et Métiers ParisTech University, Aix-en-Provence, France, in 2014, and the Ph.D. degree in electrical engineering from the University of Quebec at Trois Rivieres (UQTR), Canada, in 2018. In May 2018, he started a Postdoctoral Fellow at the Hydrogen Research Institute, UQTR. His main research interests include the optimization of energy systems for stationary and mobile applications, hybridization of energy sources for vehicular application, and eco-energy navigation of low-speed autonomous electric vehicles.