IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2, 2025

Using an Attention Mechanism for Rapid Lane Detection

[1]Dr. Abdul Rasool MD, [2]FATIMA AREEB, [3]AMENA SAMREEN, [4]THIRUMALAREDDY NIKITHA.
[1]Associate Professor, Department of AIML, Lords Institute of Engineering & Technology.
[234] Student, Department of AIML, Lords Institute of Engineering & Technology.

Abstract—

Lane recognition, a crucial subtask in autonomous driving, has recently shifted from using traditional image processing to a neural network technique based on deep learning. On the other hand, early deep learning approaches failed to match real-time needs due to their reliance on pixel-level semantic segmentation and therefore massive network architectures. A novel network architecture denoted by UFAST and built on preset rows is suggested as a solution to the real-time challenge. This network model's design achieves real-time performance by drastically reducing the network parameters. We include an attention mechanism into the model based on observations of actual human driving behaviors in order to enhance the lane recognition performance within this framework. Finally, we enhance the model framework's performance in less-than-ideal conditions by nearly 1.9% and increase the number of parameters by less than 0.2% of the UFAST. This is done by artificially scheduling the input image data to the lower part of the view, where the lanes are typically located. But actual road data is complicated, and using the same method for all the view data would result in duplicate or missing data. While the ResNet-18 [6] architecture achieved a classification accuracy of 95.87% on the TuSimple dataset [5], the CULane dataset [2] achieved just 68.4% accuracy in its trials, which is somewhat inadequate. This is because, under less-than-ideal circumstances, not only do road lane details likely to be missing from the CULane datasets because to the datasets' inherent complexity and variability, but the system framework's architecture also contributes to this problem.

Keywords—Unmanned Driving, Lane Detection, Attention Mechanisms .

INTRODUCTION

Recent years have seen tremendous growth in the unmanned driving industry, because to the ever-improving capabilities of computer vision and AI. The advancement of lane detection, a crucial component of autonomous driving, dictates the future of autonomous driving. As deep learning has evolved and improved over the last several years, it has progressively supplanted the older, more conventional lane recognition methods that relied on graphical geometry and probability distribution. Many new lane identification algorithms based on deep learning have appeared in the last few years. Much of the work in detection and classification algorithms, whether they are based on deep learning or more conventional visual information processing techniques, is going into collecting ever-more-complicated data to use as a constraint. As a result, both the total structure of detection networks and the volume of data handled have grown substantially. It is challenging to eliminate the framework complexity of pixel-level semantic segmentation, despite the fact that the newly suggested networks aim to reduce the network size. The secret to fixing the lane detecting system's sluggish performance is to eliminate unnecessary data while keeping useful data. When we look at driving patterns from a human point of view, we see that, in normal driving conditions, drivers focus on the lane they're in as well as the lanes on either side of them, from near to far, and pay less attention to the information about the boundaries of their field of vision. This procedure also serves as the system for distributing focus. What is currently missing from most lane detecting systems is a method that can simulate human lane inspection in challenging environments like poor light and vehicle obstacle. Current key attention mechanisms, from their inception, center on one or more of the following: channel[7], spatial[9], temporal[10], or branch [11] components; hybrids that mix several attention processes[12] or thirteen. The emphasis of the work is on integrating the CAMB [13] attention module with UFAST, taking into account the features of each attention mechanism and the structure of UFAST [4].

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2, 2025

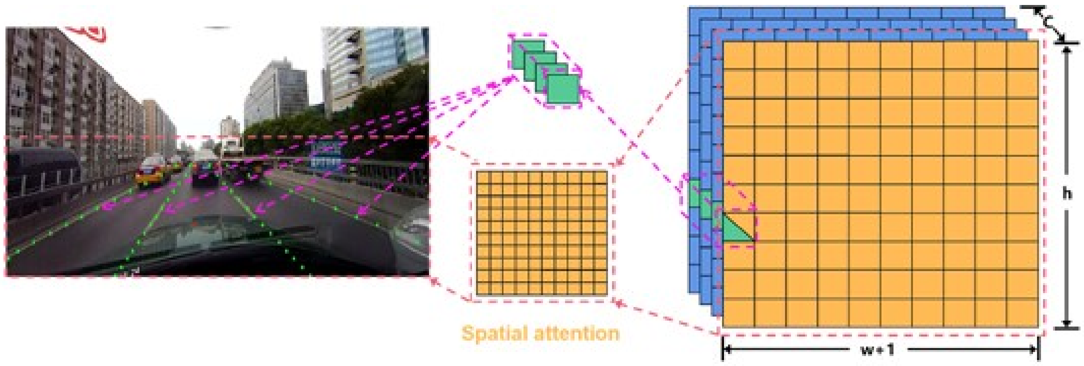

Fig.1. Rationale for the role of attentional mechanisms in the UFAST architecture.

While channel attention improves lane recognition by zeroing down on individual lane locations and shapes, spatial attention employs global feature information to help in lane detection in less-than-ideal situations, as when there is obstruction or bad lighting. For optimal system processing performance, it is necessary that the whole network learn to zero in on the lane detecting component as quickly as feasible without sacrificing processing speed. It creates a map of all the network's characteristics as well as the channels in every grid. In addition to capturing the input image's general qualities, the network also makes lane predictions in less-than-ideal situations, such as low light or crowded lanes, and partly projects these predictions into the network's overall spatial features. We use CBAM, an attention mechanism module that is very compatible with the UFAST framework, to combine these two capabilities. This component integrates channel attention with spatial attention. By incorporating the CAMB module into the entire network architecture, lane detection information may be strengthened and lane attentiveness can be improved in real-world driving scenarios. Its little stature means that it adds only one percent to the first network frame's parameter count. Our goal is to increase the UFAST framework's performance in complicated conditions, and we succeed. Here are two main elements that encapsulate our work's contribution to the field as a whole. • The system's ability to retrieve lane information more precisely thanks to the attention mechanism's implementation. • Compared to the original UFAST [4] model, the attention mechanism improves the model's accuracy in complicated road circumstances by an average of almost 1.9%. The following is the outline for this paper: Part II examines supplementary materials. The process of building the model is detailed in Section III. In Section IV, we provide not just the findings and analysis, but also the pertinent experimental setting. Our task is finally concluded in Section V.

RELATED WORK

Section A. Land Detection Techniques In the early days of lane detection, Aly et al. [14] established classic detection algorithms based on Markov random fields. Both the complexity of lane road circumstances and the speed with which deep learning has been developing in recent years have led to the proposal of deep learning-based lane recognition systems, the performance of which has been steadily improving. In their study, Davy Neven et al. [1] provide a comprehensive approach to extracting lane instances from the LaneNet network. They then use a least-squares technique to convert the lanes back to the original map, after passing the transformation matrix H produced by HNet. The LaneNet network has been an inspiration for several subsequent deep learning-based approaches. The SCNN, which was suggested in 2017 by Xingang Pan et al. [2], streamlines LaneNet's segmentation process and makes the network more succinct in general. Additionally, the CULane Datesets provide a large data set for lane recognition tests. Concurrently, Seokju Lee et al. [3] provide The VPGNet, which, by using a multi-branch job, anticipates overall lanes and zeroes down on the invisible portion of the lane prediction issue. All of the aforementioned approaches were quite accurate for their time, but the enormous network architecture of the system meant

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025

that it couldn't handle the real-time demands of autonomous driving. In 2020, Zequn Qin et al. [4] designed the UFAST network structure to address this issue. They simplified the entire detection network's backbone and used global feature pictures to creatively build a lane information model. This model significantly improved the overall detection system's speed while maintaining the state-of-the-art accuracy. For lower-speed scenarios, its real-time performance has met the needs of autonomous vehicles for lane detecting. B. Attention Mechanisms and Their Methods As a kind of attentional bias, people pay more attention to what piques their interest when they notice something. [12]. In their proposal, Volodymyr Mnih et al. [7] provide convolutional neural networks with the first RAM-implemented spatial attention. After that, SENet was suggested by Jie Hu et al. [8] as an updated model of an adaptive channel attention network; this network served as a major influence on the attention module design process for later convolutional neural networks. The CBAM attention module, proposed in 2018 by Sanghyun Woo et al. [13], combines channel and spatial attention for convolutional networks; its lightweight design allows for easy insertion between any convolutional layers with, and it is heavily inspired by SENet.

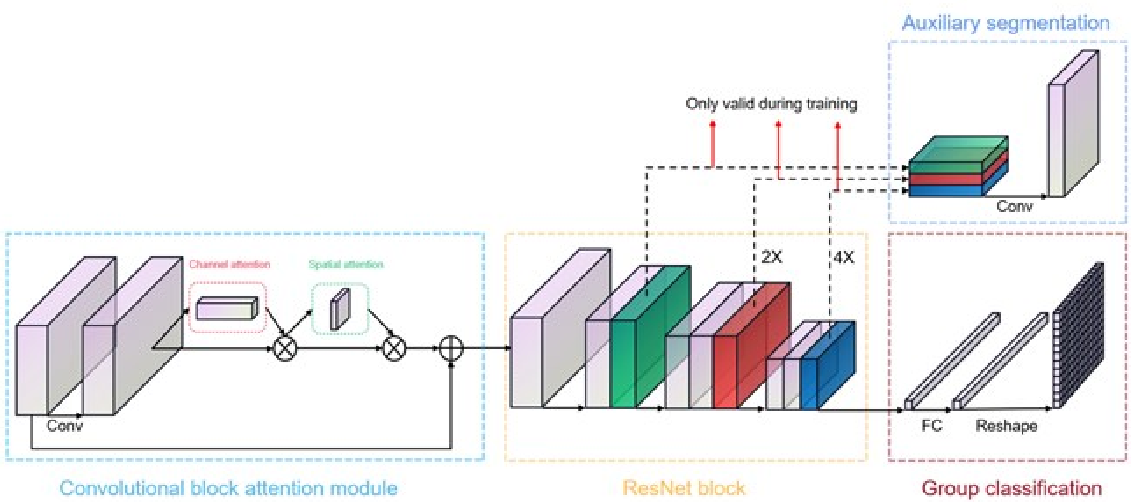

Fig. 2. Overall architecture with CBAM attention module.

The CBAM module first reads the input picture for its channel and spatial information, which it then uses to train the residual network using convolutional blocks. As shown in the top half, the auxiliary segmentation channel is formed by extracting and combining the second, third, and fourth convolutional blocks from the residual network. In order to get the final lane categorization, the convolved network is linked to the fully connected layer via the main channel. affected its network speed to a lesser extent. A number of attention mechanisms that are tailored to the architectural structure of the proposed transformer have been introduced in recent years. For example, ViT [15] has been successful in computer vision by using this structure. Nevertheless, we shall refrain from delving too deeply into the transformer design in this work.

PROCEDURE In this part, we will go over the model's implementation of the attention mechanism and the reasoning for its use to enhance network performance. Section A. Attentional Features Extraction The structural architecture that differentiates the UFAST [4] system from the semantic segmentation model is the primary cause of its large speed gain. The number of channels is proportional to the lanes in each data set, and the feed picture only intercepts the specified grid size. Consequently, compared to the semantic segmentation technique, the whole network structure is much smaller. We make the system more sensitive to lane detection by enhancing the attention to the network channel, which is the same as boosting the attention to each lane because the channel corresponds to each lane. By obtaining the picture global characteristics, UFAST is able to forecast the obstructed lanes in less-than-ideal scenarios, such as

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2, 2025

when there is not enough light or when there is a close car blocking the view. By paying more attention to the spatial characteristics of the whole convolutional network, we can better use the image's global features for lane prediction and enhance the network's feature extraction for the complete grid network. We improve the network's lane prediction and correction overall by focusing on elements like pavement margins and bridge piles. In Fig.1 we can see the application of the attention detecting overall process. Section B: Lane Detection To extract overall lanes utilizing auxiliary channels during training, we employ the ResNet-18 [6] neural network, which follows the fundamental design of UFAST [4]. Extraction of data from ResNet-18's second, third, and fourth convolutional layers, and further convolving of those layers, yields the lanes. After calibrating the lane positions in the picture and resizing to the original preset grid size, the main branch unfolds the convolutional layers to link the classifier and classifies the lanes. Later lane detection identification systems benefit from this lane attention data. The lane detecting job has been completed. While detecting in real-time, we close the secondary channel and leave the primary one open for lane identification and categorization. Fig.2 shows the final network design, which draws inspiration from the UFAST [4] and CBAM [13] networks. C. A method for detecting lanes That first UFAST loss function is still with us.

$$L_{total} = L_{cls} + \alpha L_{str} + \beta L_{seg} \quad (1)$$

TABLE I DATASETS

DESCRIPTION

Dataset	# Frame	Train	Validation	Test	Resolution	Lane	# Scenarios	environment
Tusimple	6,408	3,268	358	2,782	1280×720	≤ 5	1	highway
CULane	133,235	88,880	9,675	34,680	1640×590	≤ 4	9	Urban and highway

When the loss coefficients α and β are used, Lcls stands for the classified loss, Lstr for the structural loss, and Lseg for the segmentation loss. Following the article [4] are the specifics of these loss functions.

TABLE II NETWORK TRAINING PARAMETERS

	CULane	TuSimple
Image resize	288×800	288×800
Gridding range	260×530	160×710
Image height	540	720
Number of gridding cell	100	150
Learning rate	4e-4	4e-4
Loss coefficients(λ, α, β)	(1,1,1)	(1,1,1)
Batch size	32	32
Training epochs	50	100

Laboratory environment 1) Hardware: Instead of using the Nvidia RTX-1080Ti GPU that was used by UFAST, we are using a different Nvidia RTX-1070 GPU. This allows us to re-implement the original UFAST framework and get the following data, including its performance. 2) Datasets: TABLE I displays the features of the two datasets used to evaluate whether our model is better than the original model from UFAST [4]. We used the same benchmark datasets for lane detection as UFAST, which are the TuSimple dataset [5] and the CULane dataset [2]. 3) Measures of assessment: The two datasets are not being evaluated according to the same formal standards. Using the TuSimple dataset, the primary metrics for assessment are determined using the following formula.

$$Recall = \frac{TP}{TP+FN} \quad \text{accuracy} = \frac{\sum_{clip} C_{clip}}{\sum_{clip} S_{clip}} \quad (2)$$

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2, 2025

Sclip is the sum of all ground truth points in a clip, and Cclip is the number of lane points that were accurately predicted. This is true for the CULane dataset:

$$F1 - \text{measure} = \frac{2\times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

in this case = $Pfecibion$, $Recall$ True positives (TP), false positives (FP), and false negatives (FN) are the three possible outcomes. Displayed in TABLE II are the particulars of the image processing parameters and hyperparameters. 4)Setting up the control group: We included SENet [8] into the UFAST [4] network architecture to measure the extent to which channel attenuation alone improved network performance, with the goal of better representing the attention mechanism's contribution in enhancing overall network performance.

Results

Two features of the modified UFAST are readily apparent from the data shown in TABLES III and IV. • The model maintains the same level of accuracy as the original model on the TuSimple dataset; there is no discernible improvement. • The model improves accuracy across the board in the CULane dataset's testing domain, leading to an approximately 1.9% increase in final average accuracy when compared to the original model. The addition of the attention mechanism causes a little increase of less than 0.2 percentage points to our model parameters compared to the previous model, whereas the increase in FLOPs for convolution is around 0.4 percentage points.

TABLE III COMPARISON WITH OTHER METHODS ON TUSIMPLE TEST SET.FLOPS REFERS TO FLOATING POINT OPERATIONS IN THE CONVOLUTION . PARAM . REFERS TO THE TOTAL NUMBER OF PARAMETERS OF THE NETWORK

	UFAST-18	UFAST-18+SE	UFAST18+CBAM
Accuracy	**95.87**	95.46	95.85
FLOPs	**16.646G**	16.646G	16.653G
Param.	**44.532M**	44.619M	44.620M

TABLE IV COMPARISON OF F1-MEASURE AND RUNTIME ON CULANE TESTING SET WITH IOU THRESHOLD=0 .5 .

	UFAST-18	UFAST-18+SE	UFAST18+CBAM
Normal	87.7	88.6	**89.3**
Crowd	66	66.3	**68.3**
Night	62. 1	62.8	**64.6**
No-line	**40.2**	38.7	39.7
Shadow	62.8	58.3	**65.3**
Arrow	81	82. 1	**84.4**
Hlight	**57.9**	53.9	56.3
Curve	57.9	55.6	59.1
Total	68.4	68.5	**70.3**

Analysis We conducted the following analysis in light of the aforementioned circumstance. The original model's accuracy for the TuSimple dataset reached 95.87%, which is 2.32% different from the highest-performing SCNN UNet ConvLSTM2 [16] model and only 1.09% different from the second-highest PE RESA [17] model. This effectively reaches the optimal measurement accuracy for this dataset, and it's hard to ask for much more than that. Overall, the network over-fits and performs poorly on the test set when the attention mechanism is included. Using the CULane dataset, we discovered that, with the exception of the No-Line dataset, our model outperformed the original model in a number of more realistic scenarios. Up to a 3% improvement was the norm for the other datasets. We contend that No-Line's comparatively low model performance on this dataset is due to the attention mechanism's inability to better capture the attention object. When we apply the attention method to the model, we see that it improves the system's accuracy across the board by better capturing lane information for different conditions in the other test datasets. • The original model also does a good job of capturing lane prediction, but the TuSimple dataset is mostly used for lane recognition in a more ideal setting (light, lane line integrity, road vehicle congestion) when the two datasets are combined. On the other hand, the original model loses a lot of lane information due to the complicated and variable road conditions in the CULane dataset. However, by introducing the attention mechanism, the problem is improved and the system's accuracy is increased. This makes it more realistic and in line with how humans drive in real-life situations, when the environment is more This is also more in keeping with how people really drive in real-world situations; as things become complicated and unpredictable, drivers pay more attention to the road as a whole, which improves safety.

CONCLUSION

In order to make the lane detection model more accurate representation of how humans really pay attention to lane circumstances, we suggest including an attention mechanism in this study. We opt to enhance the UFAST [4] model and include the CBAM [13] attention mechanism module to acquire the channel and spatial properties of the lane detection model. This allows us to achieve real-time lane identification without sacrificing the original model's fast speed, which is maintained thanks to its compactness. While our improved model did not noticeably enhance lane detection in the ideal environment of the TuSimple dataset, it did generalize improve performance in multi complex cases on the CULane dataset, leading to a final average accuracy improvement of nearly 1.9%, according to our comparison experiments in the TuSimple dataset [5] and the CULane dataset [2]. It is our intention to get a better attention model design for lane recognition by enhancing the current one.

REFERENCES

[1]. [1] D. Neven, B. De Brabandere, S. Georgoulis, M. Proesmans, and L. Van Gool, "Towards end-to-end lane detection: an instance segmen- tation approach," in 2018 IEEE intelligent vehicles symposium (IV), pp. 286 291, IEEE, 2018.

[2]. [2] X. Pan, J. Shi, P. Luo, X. Wang, and X. Tang, "Spatial as deep: Spatial cnn for traffic scene understanding," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, 2018.

[3]. [3] S. Lee, J. Kim, J. Shin Yoon, S. Shin, O. Bailo, N. Kim, T.-H. Lee, H. Seok Hong, S.-H. Han, and I. So Kweon, "Vpgnet: Vanishing point guided network for lane and road marking detection and recognition," in Proceedings of the IEEE international conference on computer vision, pp. 1947– 1955, 2017.

[4]. [4] Z. Qin, H. Wang, and X. Li, "Ultra fast structure-aware deep lane detection," in European Conference on Computer Vision, pp. 276–291, Springer, 2020.

[5]. [5] 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 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 1006–1007, 2020.

[6]. [6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.

[7]. [7] V. Mnih, N. Heess, A. Graves, et al., "Recurrent models of visual attention," Advances in neural information processing systems, vol. 27, 2014.

[8]. [8] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7132–7141, 2018.

[9]. [9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł . Kaiser, and I. Polosukhin, "Attention is all you need," Advances in neural information processing systems, vol. 30, 2017.

[10]. [10] S. Xu, Y. Cheng, K. Gu, Y. Yang, S. Chang, and P. Zhou, "Jointly attentive spatial-temporal pooling network s for video-ba sed person re-identification," in Proceedings of the IEEE international conference on computer vision, pp. 4733–4742, 2017.

[11]. [11] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in Proceedings of the IEEE/CVF

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025

Conference on Computer Vision and Pattern Recognition, pp. 510–519, 2019.

[12]. [12] M.-H. Guo, T.-X. Xu, J.-J. Liu, Z.-N. Liu, P.-T. Jiang, T.-J. Mu, S.-H. Zhang, R. R. Martin, M.-M. Cheng, and S.-M. Hu, "Attention mech- anisms in computer vision: A survey," Computational Visual Media, pp. 1–38, 2022.

[13]. [13] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in Proceedings of the European conference on computer vision (ECCV), pp. 3– 19, 2018.

[14]. [14] M. Aly, "Real time detection of lane markers in urban streets," in 2008 IEEE Intelligent Vehicles Symposium, pp. 7– 12, IEEE, 2008.

[15]. [15] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, 2020.

[16]. [16] Y. Dong, S. Patil, B. van Arem, and H. Farah, "A hybrid spatial-temporal deep learning architecture for lane detection," arXiv preprint arXiv:2110.04079, 2021.

[17]. [17] J. Xie, J. Han, D. Qi, F. Chen, K. Huang, and J. Shuai, "Lane detection with position embedding," arXiv preprint arXiv:2203.12301, 2022.