ITM Web of Conferences **84**, 04015 (2026) https://doi.org/10.1051/itmconf/20268404015
ATCIDS 2026

Research and Analysis on Real-time Traffic Signal Recognition Based on Deep Learning

Weilin Chen[1]*

[1]School of Computer Engineering and Science, Shanghai University, 200444, Shanghai, China

Abstract. With the rapid development of intelligent transportation systems (ITS) and autonomous driving technology, precise and real-time traffic signal recognition has become a core technology for ensuring driving safety and improving traffic efficiency at present. However, traditional image processing methods are prone to instability in complex scenarios such as light fluctuations and adverse weather conditions, which leads to deficiencies in practical application. This paper systematically reviews the research progress of real-time traffic signal recognition technology based on deep learning: Clarify the complete evolution of this technology from traditional methods to deep learning, analyze the core challenges in the real-time recognition process and the corresponding solutions, summarize the key research results and compare the performance of mainstream technical solutions. Research and demonstration show that deep learning demonstrates a generation gap advantage in balancing accuracy and instantaneous response, establishing its position as the core evolution path of the industry. The technical paradigms and optimization details discussed in this article not only provide theoretical support for algorithm iteration, but also offer a key reference for the industrial application of intelligent traffic recognition due to its breakthrough in performance balance.

1 Introduction

The acceleration of urbanization nowadays has made the pressure on urban transportation hubs increasingly prominent. Intelligent transportation systems have become strategic battlefields for various regions to enhance efficiency and ensure safety. In the intelligent transportation systems (ITS) perception architecture, traffic signal recognition is not only the core of environmental perception but also an important part of achieving advanced autonomous driving. Through precise analysis of the semantics of traffic signals, this system can, on the one hand, help drivers avoid traffic violations, and on the other hand, rely on vehicle-road coordination technology to optimize the temporal and spatial timing sequence, thereby reducing energy efficiency loss.

Looking back at the development of technology, early recognition schemes relied on traditional computer vision paradigms. Researchers usually adopted the logic of superimposing color space transformation and morphological topological filtering to locate states [1]. However, such artificially designed geometric rules often exhibit the essential

* Corresponding author's email: weilinchen298@gmail.com

limitation of insufficient generalization when dealing with sharp fluctuations in environmental radiation brightness or global special-shaped lamps [2]. Subsequently, although machine learning algorithms represented by gradient histogram collaborative support vector machines have made breakthroughs in microscopic accuracy [3], they still cannot break away from the reliance on complex manual feature engineering, especially when locating distant and small-scale targets, the performance degradation is severe.

The development of deep neural networks has led to a paradigm shift in the field of recognition. Convolutional neural networks construct feature matrices from basic textures to deep semantics through multi-level sample refinement. At the architectural level, the two-stage detection framework represented by Faster Region-based Convolutional Neural Networks (R-CNN) has extremely high recognition confidence through the regional suggestion network [4], but its huge number of parameters limits its real-time performance on the embedded end. To break through the computing power predicament, the single-stage regression paradigm represented by the YOLO series emerged, significantly reducing inference latency [5, 6]. However, in the actual industrialization process, how to detect accuracy, throughput, find the best balance between efficiency and hardware cost is still the academia and the engineering challenges to overcome.

In addition, the complexity of the local road environment imposes extremely strict constraints on the perception algorithm. In extreme weather or low-light conditions, signal characteristics are easily swallowed up by background noise, leading to system failure [7]; Moreover, as a typical tiny target detection task, signal light pixels are prone to losing details due to the downsampling mechanism during inter-layer transmission [8]. In response to the above pain points, current research is turning to improving the loss function of bounding box regression, such as SioU [9], and attempting to introduce multimodal data fusion to construct robust stereoscopic perception barriers [10]. Meanwhile, in response to the computing power limitations of in-vehicle computing platforms, model lightweighting and hardware underlying acceleration have become new tracks for industry competition. This paper aims to systematically sort out the technical routes of real-time traffic signal recognition, deeply deconstruct core challenges such as scale sensitivity and hardware bottlenecks, and quantitatively evaluate the empirical effectiveness of mainstream optimization strategies.

2 Overview of Mainstream Technologies

2.1 Traditional Image Processing

The technical logic of traffic signal recognition has deeply evolved from the early rigid rule derivation to end-to-end semantic self-learning. Initially, the core of traditional image processing schemes lies in the binary verification of color segmentation and geometric forms. Researchers often define hard thresholds within the hue or saturation space to strip away signal regions, and then use edge detection to screen shape features [1]. Although this design based on fixed rules is highly attractive in terms of computing power cost, it is extremely fragile in real road conditions with transient light and shadow changes - light distortion often induces color drift, and this path lacks basic generalization ability for irregular lamps.

2.2 Machine Learning

Subsequently, the intervention of machine learning shifted the focus of the field to artificial feature descriptors. The typical architecture of this period relied on the Histogram of Oriented Gradients (HOG) gradient operator to depict the target contour and was combined with the linear support vector machine for classification and discrimination [3]. Although the introduction of artificial features has enhanced environmental robustness, feature engineering itself has an expressive power bottleneck. When confronted with bad weather or small targets, models often fail to capture sufficient underlying information to support classification, which can lead to significant fluctuations in recognition rates in complex scenarios.

2.3 Deep Learning

Over the past decade, the rise of deep learning has completely overturned the logic of traditional feature extraction. The technical paradigm has leaped from two-stage architectures such as Faster R-CNN, which focuses on regional suggestions and delves deeply into accuracy [4], to the YOLO series of single-stage models that pursue real-time regression and balance rate and performance [5]. By reconstructing the detection task into a single regression proposition, the YOLO framework not only maintains the recognition accuracy but also significantly reduces the inference delay, making it possible to automatically extract the compatibility of deep semantic features and high-frequency instantaneous responses.

2.4 Lightweight Model

In recent years, model lightweighting has evolved into the ultimate demand of current industrial applications. Mainstream research has found the best balance between performance and energy efficiency on embedded computing terminals such as NVIDIA by integrating the ShuffleNetV2 backbone network to achieve channel pruning and parameter quantization [6]. In particular, the introduction of spatial neck module and enhanced path aggregation network enables the model to output high-precision detection results while maintaining millisecond response, paving the way for large-scale applications of intelligent transportation [2, 6].

3 Research achievements and innovation points

This paper, through an in-depth analysis of the mainstream technological evolution paths and in combination with quantitative performance empirical evidence, summarizes a set of research results that take into account both theoretical foresight and engineering feasibility.

The evolution logic of the technical architecture indicates that single-stage detection schemes have taken the leading position in the real-time competition. The measured data show that when the traditional two-stage detector Faster R-CNN is mounted on the VGG-16 backbone network, its inference rate is only maintained at about 5 FPS. This bottleneck restricts its application in complex traffic [4]. In contrast, YOLOv3, a single-stage architecture, can successfully compress the time consumption of a single frame to 22ms through end-to-end regression, while ensuring 28.2 mAP accuracy, which provides the possibility for the deployment of embedded devices [5].

For the perception failure in today's complex environment, differentiated optimization strategy shows strong robustness. Research suggests that by implanted in YOLOv7 attention mechanism could precisely out of chaos in the background noise signal characteristics, in the weak light environment the average accuracy of 80.1%, compared to the benchmark model of 8.6% [7]. In addition, the SIoU loss function is introduced to correct the convergence offset by adding the Angle penalty factor to reshape the bounding box regression logic. This can bring 2.4% accuracy improvement, and also effectively alleviate the problem of inaccurate small target positioning [9].

This study not only overcomes the defects of monocular vision in depth perception, but also integrates the technical route from model lightweight to hardware acceleration. The unique value of this study lies in the reference of a more refined quantitative analysis framework. By analyzing the performance of modules such as SIoU, lightweight backbone network and adaptive feature fusion, the industry benchmark of real-time detection is established, and quantifiable decision support is also provided for the modular iteration of the next generation of autonomous driving sensing systems.

4 Challenges and Prospects

4.1 Challenges

4.1.1 Recognition stability under complex lighting conditions

For light upheaval and contrast of the bad weather caused, hardware and software collaborative governance solution showed stronger environmental resistance. Focus on the deployment of high dynamic range camera hardware end, with light intensity gradient in a very wide retain image detail [2]. At the algorithm level, the single recognition logic should be skipped, and the meteorological disturbance should be simulated by data enhancement, and the dynamic calibration of contrast ratio should be achieved by introducing Retinex enhancement or adaptive histogram equalization technology [2, 7]. This two-pronged strategy essentially reshapes the robustness of the system under extremely low light or strong interference conditions, so that the model is no longer trapped by variable ambient light fields.

4.1.2 The challenge of detecting small targets at distant signal lights

When the signal light is at a long distance, the pixel ratio is extremely low, and the features are prone to be lost in the downsampling of the neural network. The solution focuses on feature extraction optimization: by adopting the Feature Pyramid Network [8], deep and shallow layer features are fused through top-down and lateral connections to enhance multi-scale representation. At the same time, the attention mechanism is introduced to focus on key areas, and SIoU loss function [9] is adopted to take into account multi-dimensional geometric factors such as boundary box Angle and shape to improve positioning accuracy.

4.1.3 Real-time Requirements for Embedded Platforms

Vehicle-road coordination requires the perception system to complete single-frame processing within milliseconds. Model selection preferentially adopts single-stage detection algorithms such as the YOLO series [5,6], whose reasoning speed is superior to two-stage models such as Faster R-CNN [4]. Subsequently, the computational load was reduced through model compression techniques such as knowledge distillation, pruning and quantization [6], combined with GPU or NPU hardware acceleration, real-time response was achieved on edge platforms such as Jetson AGX Orin.

4.2 Prospects

Monocular visual depth lack for the shortcomings, further research should be realized in perception, algorithm and system three dimensions deep breakthrough. At the hardware level, RGB-D sensors are introduced to obtain three-dimensional point clouds, and complex backgrounds are stripped through high-dimensional information to solve the recognition problems of occluded and overlapping targets [10]. Algorithmically, it is necessary to break away from the path dependence on high-cost labeled data and shift to the self-supervised learning paradigm, using pre-tasks such as image restoration to drive the model to capture the intrinsic features of the scene. The system architecture should abandon the traditional module stacking and try to fuse multi-modal data such as millimeter wave radar through transformers to build an end-to-end unified perception matrix [10]. Such a leap from local recognition to global semantic analysis will greatly reduce the accumulation of errors and give the system robust migration capability under complex operating conditions.

5 Conclusion

This paper explores real-time traffic signal recognition technology based on deep learning and clarifies the complete technological evolution from traditional image processing to deep learning lightweight models. In response to the three core technical pain points of sudden illumination changes, small target detection, and real-time constraints, a complete set of software and hardware collaborative optimization solutions has been formed. Research has confirmed that the existing technologies have a solid foundation and can meet the basic perception requirements of intelligent transportation and autonomous driving. In the future, continuous breakthroughs can be made in three aspects: perception hardware, algorithm architecture and system design. The reliability and generalization of this technology will be further enhanced, providing key technical support for the construction of intelligent transportation systems and the industrialization of autonomous driving. It has both significant theoretical value and engineering application significance.

References

[1] X. Chen, Y. Chen, and G. Zhang. A computer vision algorithm for locating and recognizing traffic signal control light status and countdown time. J. Intell. Transp. Syst. 25, 533–546 (2021). DOI: 10.1080/15472450.2021.1871611
[2] X. Zhang, Y. Feng, P. Angeloudis, and Y. Demiris. Monocular Visual Traffic Surveillance: A Review. IEEE Trans. Intell. Transp. Syst. 23, 14148-14165 (2022). DOI: 10.1109/TITS.2022.3147770
[3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). San Diego, CA, USA. (2005). 886-893
[4] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans. Pattern Anal. Mach. Intell. 39, 1137-1149 (2017). DOI: 10.1109/TPAMI.2016.2577031
[5] J. Redmon and A. Farhadi. Yolov3: An incremental improvement. arXiv Prepr. (2018).
[6] Q. Liu, Z. Zhao, J. Wang, and S. Pi. High-performance YOLOv5: Research on High-Performance Object Detection Algorithm for Embedded Platforms. J. Electron. Inf. Technol. 45, 2205-2215 (2023). DOI: 10.11999/JEIT220413
[7] D. Zhao, F. Shao, S. Zhang, L. Yang, H. Zhang, S. Liu, and Q. Liu. Advanced Object Detection in Low-Light Conditions: Enhancements to YOLOv7 Framework. Remote Sens. 16, 4493 (2024). DOI: 10.3390/rs16234493
[8] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie. Feature Pyramid Networks for Object Detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA. (2017). 936-944
[9] Z. Gevorgyan. SIoU loss: More powerful learning for bounding box regression. arXiv Prepr. (2022).
[10] D. Fang. Research on the Application of 3D Object Detection Technology Based on Modal Fusion in Intelligent Transportation Systems. (Shanghai University, Shanghai, 2023).