Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/5/20230586

# Current research and progress of machine learning in traffic monitoring systems

**Zhengqian Wang**

Information Technology, Shanghai Ocean University, Shanghai, China, 200070

shimenlu@outlook.com

**Abstract.** In modern traffic environments and road construction, road monitoring is one of the most important safeguards for urban safety. Traditional image processing techniques are not suitable for comprehensive data collection and thus accurate and fast feature extraction and generalization of the massive amount of data generated. Real-time surveillance is both tedious and tiring, but vital to safety. This paper describes the traditional image and video processing techniques that have been used to monitor pavements, as well as today's means of image processing for pavement inspection, and summarize the current development based on existing literature research. The result shows that the application of these methods can lead to a better understanding of complex pavement conditions, improved road planning and design and overall safety, and even derive applications to other aspects of computer vision as well.

**Keywords:** machine learning, image processing, traffic monitoring, traffic density, computer vision.

## 1. Introduction

As the world's population grows, the increased density of traffic leads to an increasing amount of road congestion, air pollution and traffic accidents. The total number of commutes worldwide has increased exponentially in the last decade. In this context, road monitoring is undoubtedly a major challenge for many Chinese cities. The rage of the new crown epidemic has brought about a surge in demand for issues such as flow regulation. Traditional means of collecting and acquiring information have long been unable to cope with today's ever-expanding volume of data, but the sheer volume of data has brought another benefit, big data, which was previously unavailable. On the other hand, the exponential rise in costs and concerns over privacy and security caused by the installation and maintenance of unlimited, high-volume surveillance equipment across the country has called for the development of a "fewer but better" strategy, with fine-tuned monitoring and management of selected key roads[1]. In contrast, some of the traditional image processing methods still in use today are more or less unsuited to the productive needs of today's life. This has led to the creation of a system. This paper introduces the traditional image and video processing technology used to monitor the road surface as well as the image processing means used for road surface detection today. Finally, some assumptions are put forward for promising development in the future. The research presented in this paper helps enrich the research on this topic and lay a foundation for further in-depth research on this topic in the future.

DOI: 10.54254/2755-2721/5/20230586

## 2. Traditional image processing techniques

Image technology is one of the computer technologies commonly used in life. It plays an irreplaceable role in many application technology fields, such as telecommunications systems, automation systems in autonomous vehicles, monitoring systems and medical research fields. Traditional image processing methods generally refer to a series of operations such as segmentation, transformation, compression, filtering, restoration, and reconstruction as well as recognition and feature extraction of images using various methods of image filtering and orthogonal transformation[2]. Traditional image processing tasks were different from those of today. In the early stage of image processing, the domestic purpose was to improve the image quality to achieve the purpose of improving the visual effect[3]. In this process, the pictures with poor visual effects are input to get more favorable pictures for people to watch. General digital image processing includes image enhancement, restoration, compression, coding, etc. Later on, image processing techniques such as geometric correction, gradient transformation, and noise cancellation were also gradually introduced and became familiar to us as traditional image processing. Today, as general-purpose computers have become faster and faster, they have almost completely taken over the role of dedicated hardware outside of some of the most specialized niche areas. Additionally, digital image processing has emerged as the most widely utilized method of image processing and a result of the increasing use of computers and signal processors. Because of its affordability as well as versatility and ease.

## 3. Transition of traditional image processing methods to new technologies

Research topics that used to be the result of small datasets available for training have been opened up to technological progress by the availability of scaled-up models. For instance, deep neural networks trained on enormous data sets have lately supplanted traditional methods based on hand-crafted features and standard machine learning techniques in other computer vision topics, such as face recognition. The methods used for face recognition have greatly evolved throughout time. Traditional approaches use machine learning methods like principal component analysis, linear discriminant analysis, or support vector machines in conjunction with hand-crafted features like edge and texture descriptors[4]. However, the processing method of handmade features will produce a phenomenon called robustness in an unrestricted test environment[5]. Therefore, researchers have to deal with different situations and provide different processing methods, such as lighting, posture, age and so on. Deep learning techniques based on convolutional neural networks(CNN) have lately supplanted conventional face recognition techniques[6]. The advantage of deep learning is that it can use a large number of data sets for training, so as to obtain the characteristics that meet the requirements. The data set of face training can be conveniently collected directly from the Internet, and these faces cover almost all features, which may accurately reflect the features of the face in the actual world. Using these datasets for training, we can get the face recognition results with robust features in the real world, which is also the advantage of CNN based face recognition methods compared with the past manual methods[7]. In this respect, we can also draw similar conclusions from the road detection, and the development of in-depth learning is accelerating the progress of related industries.

## 4. Improved image and video processing

Among the road detection and analysis methods that have been widely used, the typical machine learning methods are ANN, SVM, CNN, etc. These methods are used to detect the degree of pavement disease. The CNN stated above is an example of a common neural network that computes via convolution[8]. CNN naturally outperform ANN in terms of computing efficiency since they can use more layers than ANN, which are limited to fully connected ones. The convolutional kernel's shared parameters and the layers' local connectivity enable it to carry out repeated feature learning tasks at a lower computing cost[9]. For the same number of layers, the number of weights is significantly less than that of an ANN. Convolutional, pooling, and fully connected layers, to mention a few, are only a few of the functional layers that make up CNN's hidden layers in contrast to ANN[9].

DOI: 10.54254/2755-2721/5/20230586

The convolution layer's job is to convolve the data input. The pooling layer's job is to choose and filter the data that was retrieved from the convolution layer, which results in a smaller model, faster computation, and improved resilience of the extracted features. The size, collapse, and padding of the filter are the hyperparameters[10]. Maximum pooling and average pooling techniques are typically employed. While the average pooling approach uses the average of the output pooled region, the maximum pooling method employs the largest value of the pooled region as the new feature output[10]. The pooling layers are typically placed behind several succeeding convolutional layers in deep CNN structures, whereas numerous fully connected layers are placed at the end of the entire network. The completely linked layers are comparable to a typical neural network's hidden layers. The 3D structure is lost, extended into a vector, and passed into the following layer via the activation function after the feature map has been transmitted into the fully-connected layer. The most active area of research in recent years has been the selection of loss function for training CNN-based techniques. Despite the great success of CNNs trained with softmax loss, it has been claimed that the usage of such loss functions does not generalize well to objects that are not part of the training set. This is because softmax loss is encouraged to learn features that don't necessarily reduce within-class variance but instead enhance between-class variation(to be able to distinguish classes in the training set). Many solutions have been put forth to solve this issue. Utilizing joint Bayesian or other discriminative subspace methods to optimize bottleneck characteristics is a straightforward strategy[11]. Of course, metric learning is also used in another strategy.

## 5. Current status of traffic system monitoring models

Video surveillance and image processing technology are widely used in traffic management. Video monitoring can manage the density of traffic flow, distinguish the types of vehicles, and detect the occurrence of traffic accidents. Use the monitoring camera installed at the intersection to record, capture and analyze the traffic conditions on the road in real time. Based on the collected data, you can adjust the interval of signal lights on congested roads in real time, greatly improve the commuting efficiency, reduce the occurrence of traffic accidents, and save time and fuel consumption for all commuters. Moreover, these data can also be used to analyze traffic conditions, commuter habits, etc., which is very helpful for improving urban road construction and planning infrastructure. Relevant technologies can also be used to analyze traffic conditions, guide traffic flow, statistics pedestrian information, traffic safety, pavement quality detection, license plate number recognition, accident analysis, etc. Even though road sensors installed on both sides of the road or buried underground are considered worthy of investment due to their sensitivity and low purchase cost, compared with surveillance cameras, they require closed construction of relevant roads, and are more difficult and costly to install, providing limited information.

Here are some schemes proposed by predecessors to demonstrate the feasibility of the proposed design of this system. A traffic monitoring was approached by Sullivan, Tan, and colleagues as a model-based tracking problem, and a complete 3D analysis was carried out to gather vehicle information[12]. They employed a wireframe model of the vehicle, which was back-projected into the image based on the estimated pose(position and orientation), and comprehensive calibration information. The model's edges were then matched to the lines in the image. Koller et al. tried to establish a system to monitor the crossing. They used a three-dimensional model and preset the model frame to match the cut image. When the system starts to run, the system will make an estimate based on the current position and put it back into the image to calculate the new position[13]. Vandenberghe et al. proposed the feasibility of extending the traffic detection system using floating car data techniques[14]. The location of traffic incidents, the end of traffic jams on all sorts of roads, and the amount of time needed to traverse each segment of road may all be accurately identified using moving automobile data to enhance traffic monitoring. SiamFC was first proposed by Bertinetto et al. using the deep learning based tracker ConvNets[15]. This tracker consists of two branches that take two inputs — a search image and an example — and apply the identical alteration to both. Then, using cross correlation, the two representations are joined to produce a fractional graph showing the object's most likely location. Li et al. proposed SiamRPN, adding the regional proposal network to the conjoined network to generate the boundary box proposal after classification and regression branches[16]. The vehicle detection method in the tunnel was proposed by Huang et al.[17]. To reduce the interference of the special background of the tunnel, special cameras were used in the tunnel and erected in different ways. At the same time, background subtraction and deep confidence network are used to solve the problems of insufficient light source of camera results, blurry imaging, difficult resolution and many interfering objects. The video surveillance system used by Ukani to detect and classify vehicles[18]. The technologies used are support vector machine and neural network. The size invariant transformation features are extracted from the detection results. The results show that SVM has better generalization ability than artificial neural network.

## 6. Conclusion

This paper presents its views on the construction of road detection systems based on theories related to digital image technology and deep learning frameworks, as well as on the practical operation of video image processing technology. The construction of improved road detection systems is of great importance to the development of our urban life. Image processing methods for road detection began the transition, as did many other computer vision applications. A few years ago, the traditional method based on manually designed features could provide reliable accuracy, but now it is almost completely replaced by the CNN based deep learning method. In fact, compared with other types of methods, CNN based face recognition system has indeed achieved significant accuracy improvement. In addition, with the increasing amount of information, the training set also expands, and it is easier to achieve higher accuracy through deep learning neural network training. At the same time, the system can also be assisted by other data collection means mentioned above to enhance the overall effect and make up for the slow training and deployment of CNN deep learning architecture. The application of computer vision in road detection can also be developed in other areas such as human behavioral observation, animal activity observation and detection, etc. After strengthening algorithms and computing power, it can be used for detailed aspects such as medical human bodies, metal welders' welding joints, antique restoration, etc. The spread of deep learning methods could also accelerate the development of computer vision. The current research in this thesis summarizes the various video image tracking techniques that have been proposed by previous authors. Future research will conduct in-depth research on methods such as deep learning for feature extraction at a later date, and to compare the results and make an intuitive comparison.

## References

[1] Rusu, L. , Taut, D. , & Jecan, S. (2015). An integrated solution for pavement management and monitoring systems. Procedia Economics and Finance, 27, 14-21.

[2] B. Basavaprasad and S. H. Ravindra. 2014. A survey on traditional and graph theoretical techniques for image segmentation. International Journal of Computer Applications, 975:8887.

[3] R. Chandel and G. Gupta. 2013. Image filtering algorithms and techniques: A review. International Journal of Advanced Research in Computer Science and Software Engineering, 3(10), 2

[4] Jiang, E. . (2020). A review of the comparative studies on traditional and intelligent face recognition methods. 2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL).

[5] Amirshahi, S. A., Pedersen, M., & Beghdadi, A. (2018, November). Reviving traditional image quality metrics using CNNs. In Color and imaging conference (Vol. 2018, No. 1, pp. 241-246). Society for Imaging Science and Technology.

[6] Trigueros D S, Meng L, Hartnett M. (2018). Face recognition: From traditional to deep learning methods[J]. arXiv preprint arXiv:1811.00116.

[7] Li, C., Ma, P., Rahaman, M. M., Yao, Y., Zhang, J., Zou, S., ... & Grzegorzek, M. (2021). A State-of-the-art Survey of Object Detection Techniques in Microorganism Image Analysis:

DOI: 10.54254/2755-2721/5/20230586

from Traditional Image Processing and Classical Machine Learning to Current Deep Convolutional Neural Networks and Potential Visual Transformers. arXiv preprint arXiv:2105.03148.
[8] Hou Y, Li Q, Zhang C, et al. (2021). The state-of-the-art review on applications of intrusive sensing, image processing techniques, and machine learning methods in pavement monitoring and analysis[J]. Engineering, 7(6): 845-856.
[9] A. Limshuebchuey, R. Duangsoithong, and M. Saejia. (2020). Comparison of image denoising using traditional filter and deep learning methods. In 2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), pages 193–196. IEEE.
[10] Y. Hou, Q. Li, C. Zhang, G. Lu, Z. Ye, Y. Chen, L. Wang, and D. Cao. (2021). The state-of-the-art review on applications of intrusive sensing, image processing techniques, and machine learning methods in pavement monitoring and analysis. Engineering, 7(6):845–856.
[11] Taha, M. A., Şah, M., & Direkoğlu, C. (2020, August). Review of place recognition approaches: traditional and deep learning methods. In International Conference on Theory and Applications of Fuzzy Systems and Soft Computing (pp. 183-191). Springer, Cham.
[12] Sullivan, G. D. , & Sloman, A. (1992). Visual interpretation of known objects in constrained scenes[and discussion]. Philosophical Transactions of the Royal Society B: Biological Sciences.
[13] D., Daniilidis, K., Thorhallson, T., & Nagel, H. H. (1992, May). Model-based object tracking in traffic scenes. In European Conference on Computer Vision (pp. 437-452). Springer, Berlin, Heidelberg.
[14] Vandenberghe W, Vanhauwaert E, Verbrugge S, et al. Feasibility of expanding traffic monitoring systems with floating car data technology[J]. IET Intelligent Transport Systems, 2012, 6(4): 347-354.
[15] Bertinetto, L., Valmadre, J., Henriques, J. F., Vedaldi, A., & Torr, P. H. (2016, October). Fully-convolutional siamese networks for object tracking. In European conference on computer vision (pp. 850-865). Springer, Cham.
[16] Li, B., Yan, J., Wu, W., Zhu, Z., & Hu, X. (2018). High performance visual tracking with siamese region proposal network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8971-8980).
[17] Huang, B. J., Hsieh, J. W., & Tsai, C. M. (2017, April). Vehicle detection in Hsuehshan tunnel using background subtraction and deep belief network. In Asian Conference on Intelligent Information and Database Systems (pp. 217-226). Springer, Cham.
[18] Ukani, V. G. (2017). Efficient vehicle detection and classification for traffic surveillance system. International Conference on Advances in Computing and Data Sciences, pp. 495–503.