ABSTRACT

of the dissertation work by Olzhas Muratuly Olzhayev titled «Development of a road damage detection system using deep learning methods based on video data», submitted for the degree of Doctor of Philosophy (PhD) in the educational program 8D06105 - «Data Science»

Relevance

Ensuring the operational reliability of road infrastructure should be considered a critical factor determining not only the pace of economic development but also the level of public safety. Road surface deterioration typically begins with the formation of a network of microcracks, which, if not addressed promptly, inevitably develop into deep potholes. Such defects pose a direct threat to road users. According to global reports from the World Health Organization (WHO), annual deaths from road accidents reach 1.19 million. Meanwhile, non-fatal injuries, often resulting in permanent disability, account for up to 50 million victims. It is important to consider that road surface condition accounts for at least 10% of all recorded accidents. Therefore, the transition to road network monitoring is transforming from a narrowly technical task into a priority public health strategy.

Direct costs of road accident repair directly impact the financial impact of operating deteriorated road networks. Global research shows that the combined costs of road traffic injuries and road surface degradation can reach 3% of a country's gross domestic product (GDP). This figure is driven by direct medical costs, long-term declines in labor productivity, and accelerated wear and tear on engineering structures and vehicles. Furthermore, billions are wasted annually on unscheduled maintenance due to vehicle fleets operating on defective surfaces, leading to premature exhaustion of their technical resources. Given these circumstances, as road agencies strive to optimize operating costs, the transition to proactive monitoring strategies becomes the only alternative to asset maintenance.

Visual inspections delegated to maintenance personnel remain a popular method of road network auditing. However, the potential depth of such expert assessment is mitigated by limited technical resources. In reality, manual inspection is an extremely inert process, associated with high operational costs and, critical for data accuracy, a significant risk of subjective bias. The physical limitations of human error make it impossible to ensure adequate spatial coverage necessary for continuous monitoring of road networks in real time. Against the backdrop of rapid urbanization and the increasing complexity of transport hub topologies, traditional visual inspection has ultimately lost its effectiveness and is hindering the development of a digital road maintenance environment.

The rapid evolution of computer vision and deep learning algorithms has opened up fundamentally new possibilities for the creation of autonomous road infrastructure monitoring systems. Breakthrough results achieved in classification, object detection, and semantic segmentation have focused research interest on automating the detection of road defects. Within this technological vector, the use of deep convolutional neural networks (CNNs) and modern detectors has become the dominant approach. Such architectures enable the extraction of high-level visual features from high-resolution images, ensuring precise damage identification in complex visual contexts.

Automated road defect detection using computer vision is currently considered a hot topic, and interest in this field is growing exponentially. An analysis of relevant literature clearly demonstrates this surge in scientific activity: while in 2015 there were only about 50 relevant publications, by 2025 their number exceeded 900. This growth is driven by three factors: the widespread adoption of high-resolution sensors, the development of large-scale representative datasets, and the availability of high-performance computing power necessary for training multilayer neural networks.

If we analyze current scientific research, we notice that they primarily focus on processing single images from dashcams or smartphones. However, practice shows that this is insufficient for real-world monitoring. We need to see the road surface condition dynamically using a video stream. Working with video adds a "time coordinate" to the system, which helps more accurately identify defects and filter out random interference. When a vehicle's camera records video continuously while driving, it allows for coverage of vast sections of the road network. The result is a tool for diagnosing damage in near real time, which is critical for prompt highway maintenance.

Despite the progress made, automated road defect detection using remote sensing still faces a number of fundamental barriers. The effectiveness of existing systems is critically dependent on environmental variability: the dynamic range of illumination, the presence of deep occlusions (shadows), seasonal landscape changes, and unstable weather conditions. Furthermore, the complex nature of remote imaging at acute angles must be taken into account, which inevitably leads to perspective distortions of objects. Moreover, the morphology of the defects themselves—their texture, nonlinear shapes, and wide range of geometric dimensions—makes their precise differentiation extremely difficult using standard methods. These challenges necessitate a transition to more complex, hierarchical deep learning architectures. In particular, this requires the implementation of specialized feature extraction modules that are highly robust to visual noise and capable of identifying relevant damage patterns in low-contrast scenes.

The main **goal** of the dissertation research is to develop an intelligent system for automatic detection and analysis of road surface damage in real time based on computer vision and deep learning methods, ensuring increased accuracy in identifying road infrastructure defects and the possibility of their subsequent mapping.

The following research **objectives** were set:

- collect video data using high-resolution cameras and other recording devices.
- perform manual data labeling, identifying various types of road surface damage, such as cracks, potholes, and deformations.
- create a high-quality labeled dataset for training and testing deep learning models.
- analyze visual features and characteristics of road surface images that affect data quality and the robustness of deep learning models.
- develop and train a deep learning model for road damage classification and segmentation.
- test the developed models on the collected dataset to achieve high accuracy and robustness to various shooting conditions.
- integrate the trained model into an application for automatic road condition monitoring, enabling the creation of interactive damage maps.

The **object** of the research is the transport and operational condition of highways and the visual characteristics of road surface defects.

The **subject** of the research is methods and algorithms of computer vision, deep learning and multi-task neural networks for detection, classification and segmentation of road surface damage in real time.

Methodology

The methodological basis of the work is based on digital image processing methods, computer vision algorithms, and modern deep machine learning techniques. The study utilizes convolutional neural networks (CNNs), transformer attention mechanisms, multi-task learning, and video data analysis methods to build an intelligent road damage detection system. Methods of data preprocessing, image augmentation, neural network optimization, and experimental evaluation of model quality using detection and segmentation metrics are also employed.

Scientific provisions submitted for defense:

- a developed multi-task neural network architecture (TCR-RoadNet) for simultaneous detection, classification, and segmentation of road surface damage in real time.
- a developed contextual feature refinement module based on cross-scale transformer attention, improving the localization accuracy of complex and fragmented defects.
- a proposed method for generating invariant visual features, ensuring defect recognition robustness to changes in illumination, background noise, and weather conditions.

Main results of the research:

- an intelligent system for automatic road surface damage detection based on video data analysis was developed, enabling continuous monitoring of road infrastructure in real time using computer vision and deep learning methods.
- a multi-task neural network architecture, TCR-RoadNet, was proposed for the simultaneous detection, classification, and segmentation of road defects. The architecture includes a multi-scale convolutional feature extraction unit, a Transformer Context Refinement (TCR) module, and specialized processing branches, which improved the localization accuracy and robustness of complex damage recognition.
- a contextual feature refinement module based on a transformer attention mechanism was developed and implemented, enabling effective analysis of spatial dependencies between objects of different scales. This module improved segmentation quality and reduced the number of false positives in the presence of shadows, glare, and uneven asphalt texture.
- a proprietary dataset of road surface video data containing various types of damage, including longitudinal and transverse cracks, mesh damage, and potholes, recorded in real-world transport infrastructure operating conditions, was collected, pre-processed, and labeled.
- the effectiveness of the developed model was experimentally evaluated. The results of computational experiments demonstrated stable convergence of the neural network during training, high defect detection and segmentation accuracy, and the system's ability to operate in real time at up to 57 frames per second.
- a comparative analysis of the developed architecture with existing deep learning models for road monitoring tasks was performed, demonstrating improved defect detection accuracy and segmentation quality while maintaining high system performance.

- software and a web interface for intelligent road infrastructure monitoring were developed, enabling automatic visualization of detected damage, georeferencing, and display of analysis results on an interactive map for further use by road services and utility organizations.

Scientific novelty

The scientific novelty lies in the development of a multi-task neural network model, TCR-RoadNet, which simultaneously detects, recognizes, and segments road defects in real time. A key element of this innovation is the implementation of a transformer attention mechanism, which enables the analysis of global spatial dependencies of the road scene. This enabled the development of a unique adaptation method, enabling the system to reliably detect defects in any weather and lighting conditions, ignoring shadows, puddles, and visual noise.

Scientific contribution:

- a unique dataset of road surface images containing various types of damage recorded in real-world transport infrastructure operating conditions was developed and labeled.
- additional visual features, such as the influence of seasonal conditions and dynamic lighting variability, were identified and algorithmically taken into account, significantly increasing the model's robustness and resilience to challenging shooting conditions.
- a comprehensive automated road infrastructure monitoring system was developed, capable of processing continuous video streams in hard real-time and providing analytical reporting via an integrated web interface.

Theoretical significance

The theoretical significance of this work lies in the application of computer vision and deep learning methods to the problem of continuous road monitoring and analysis. Unlike most current approaches that process static images, this work expands the scientific and methodological basis for analyzing the spatiotemporal characteristics of video sequences. The developed methods facilitate algorithmic support for intelligent systems for the development of reliable machine vision systems in complex environments.

Practical significance

The practical significance of this research lies in the creation of ready-to-implement software and a specialized information system that automates the process of detecting road defects and facilitates the transition from manual inspections to intelligent road infrastructure monitoring systems. The study resulted in a studied and validated neural network model for classifying and localizing road defects (potholes, cracks) based on real-time video monitoring equipment, as well as a specialized web application implemented as an interactive control panel. Integration of the trained neural network model with web interface data allows for the automatic mapping of detected road defects and their display on a map. The obtained results and the developed software can be used for the rapid response of road services, for the objective determination of the priority order of road repairs and for the prevention of the expenditure of the budget on technical maintenance by specialized public utilities, road construction companies and the ministries of transport and road infrastructure.

Reliability of results

The reliability of the obtained results is ensured using modern computer vision and deep learning methods, the application of convolutional neural networks (CNNs) and transformer attention mechanisms, as well as computational experiments on a labeled dataset of road surface video collected under real-world operating conditions. The reliability of the results is confirmed by a comparative analysis with existing architectures and algorithms for road defect detection, the use of generally accepted quality assessment metrics (mAP, IoU, Precision, Recall, F1-score), stable model convergence during training, and the reproducibility of the obtained results under various lighting conditions, weather factors, and background noise.

Approbation of the dissertation work

The main results of the work were presented and discussed at seminars of the Department of Mathematical and Computer Modeling of JSC IITU (2024–2026) and Asia Metropolitan University, Malaysia (2026). Four publications on the topic of the dissertation have been published, including two in top-ranked scientific journals indexed in the Scopus database; two publications in the proceedings of international conferences; and one author's certificate has been received.

1. Olzhayev, O. M., Kulambayev, B. O., Sakenkyzy, N., & Belisbek, M. (2026). A Real-Time Multi-Scale Feature Pyramid YOLO Architecture for Accurate and Deployment-Efficient Road Damage Detection. *International Journal of Advanced Computer Science & Applications*, 17(3). https://doi.org/10.14569/IJACSA.2026.0170350

2. Kulambayev, B. O., Olzhayev, O. M., Altayeva, A. B., & Zhunisbekova, Z. (2025). A Multi-Scale ROI-Aligned Deep Learning Framework for Automated Road Damage Detection and Severity Assessment. *International Journal of Advanced Computer Science & Applications*, *16*(12). https://doi.org/10.14569/IJACSA.2025.01612107

3. Olzhayev, O., Kulambayev, B., & Omarov, B. (2025). Real-Time Pixel-Wise Segmentation of Road Surface Damage Using a 2D U-Net Architecture. *Procedia Computer Science*, *269*, 131-139. https://doi.org/10.1016/j.procs.2025.08.266

4. Kulambayev, B., & Olzhayev, O. (2025). A Mask R-CNN Algorithm for Automated Segmentation of Asphalt Road Cracks. *Procedia Computer Science*, *269*, 39-48. https://doi.org/10.1016/j.procs.2025.08.257

5. Olzhayev O. Certificate of the Right to Protect a Computer Program No. 69666 of the Republic of Kazakhstan. TCR-RoadNet Software, dated April 7, 2026.

In all the listed publications, the applicant plays a leading role, and the main research results, analyses, models, and programs are created by the author, and the conclusions are based on the results obtained from the applicant's work and research.

Connection with state programs

This dissertation was funded by grant funding (IRN AP23487192) on the topic "Development of a Real-Time Road Damage Detection System Using Computer Vision and Artificial Intelligence".

This research aligns with the strategic development priorities of the Republic of Kazakhstan and contributes to the implementation of the Concept for the Development of Artificial Intelligence in the Republic of Kazakhstan for 2024–2029, approved by Government Resolution No. 592 of July 24, 2024, in terms of the development and implementation of intelligent computer vision systems, video data analysis, and automated monitoring of transport infrastructure. The developed multitasking neural network architecture, TCR-RoadNet, and the road damage detection system contribute to the development of domestic artificial intelligence technologies, intelligent transport systems, and the digitalization of Kazakhstan's road infrastructure.

Main content of dissertation

The dissertation focuses on the problem of defect detection, classification and segmentation of road surface damage based on video data analysis, using advanced deep learning methods.

The first section substantiates the relevance of the problem of road infrastructure degradation and the need to transition from resource-intensive manual inspections to intelligent monitoring systems. The goal and objectives of the study are formulated, and the scientific novelty and practical significance of the developed multitasking neural network architecture and software and analytical suite are determined.

The second section provides a comprehensive literature review, including a systematization of road defect types and an analysis of the evolution of detection methods. The historical transition from classical computer vision algorithms to modern deep learning architectures (convolutional networks, YOLO detectors, visual transformers) is examined. Fundamental gaps in existing research are identified, such as the difficulty of processing continuous video streams due to temporal inconsistency, the lack of representative video datasets, and the challenge of adapting heavyweight models to edge computing, which algorithmically justifies the current research direction.

The third section describes the research materials and methods in detail. The process of collecting, preprocessing, and precision annotation of a unique video dataset captured in real-world conditions is presented. The focus is on the mathematical and structural design of the innovative multi-task neural network architecture, TCR-RoadNet. The operation of the multi-scale convolutional backbone, the contextual refinement module based on cross-scale transformer attention (TCR), and three specialized inference branches: a separate detection module, a classification refinement block, and an edge-aware segmentation branch are described in detail. The optimization strategy and multi-task loss functions (a combination of CIoU, Focal Loss, and BCE+Dice) are also presented.

Section four presents the results of a comprehensive computational and experimental evaluation of the developed system. The model's convergence dynamics during training are analyzed, and semantic accuracy is quantitatively assessed using the standardized mAP and mIoU metrics. A detailed ablation study is conducted, empirically proving the architectural feasibility of each implemented computing module. A comparative analysis with state-of-the-art global peers confirms that the proposed TCR-RoadNet model provides an optimal compromise between high-accuracy, fine-grained recognition, and the ability to operate in hard real-time at 57 frames per second. A qualitative analysis demonstrates the network's ability to ignore background noise and shadows.

Section five summarizes the overall findings of the dissertation. The successful achievement of the stated goal is confirmed, the main scientific results are summarized, and conclusions are formulated regarding the readiness of the developed hardware and software complex for practical implementation in the tasks of automated audit of transport networks.

The total volume of the dissertation is 104 pages, including 28 illustrations, 8 tables and a list of 154 references.