Real-Time Road Surface Damage Detection Framework based on Mask R-CNN Model

Bakhytzhan Kulambayev[1], Magzat Nurlybek[2], Gulnar Astaubayeva[3], Gulnara Tleuberdiyeva[4], Serik Zholdasbayev[5], Abdimukhan Tolep[6]

Turan University, Almaty, Kazakhstan[1]
Bachelor Student at Turan University, Almaty, Kazakhstan[2]
NARXOZ University, Almaty, Kazakhstan[3, 4]
International Information Technology University, Almaty, Kazakhstan[5]
Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkistan, Kazakhstan[6]

***Abstract*—In the ever-evolving realm of infrastructure management, the timely and accurate detection of road surface damages is imperative for the longevity and safety of transportation networks. This research paper introduces a pioneering framework centered on the Mask R-CNN (Region-based Convolutional Neural Networks) model for real-time road surface damage detection. The overarching methodology encapsulates a deep learning-based approach to discern and classify various road aberrations such as potholes, cracks, and rutting. The chosen Mask R-CNN architecture, renowned for its proficiency in instance segmentation tasks, has been fine-tuned and optimized specifically for the unique challenges posed by road surfaces under diverse lighting and environmental conditions. A diverse dataset, amalgamating urban, suburban, and rural roadways under varied climatic conditions, served as the foundation for model training and validation. Preliminary results have not only underscored the model's robustness in real-time detection but also its superiority in terms of accuracy and computational efficiency when juxtaposed with extant methods. Concomitantly, the framework emphasizes scalability and adaptability, positing it as a frontrunner for potential integration into automated road maintenance systems and vehicular navigation aids. This trailblazing endeavor elucidates the potentialities of deep learning paradigms in revolutionizing road management systems, thus fostering safer and more efficient transportation environments.**

Keywords—Deep learning; CNN; random forest; SVM; neural network; prediction; analysis

I. Introduction

Road infrastructure remains a pivotal element in the socio-economic fabric of nations, serving as the backbone of trade, transportation, and daily commuting [1]. As urbanization and globalization continue to expand, so does the reliance on a durable and well-maintained road network. While the necessity of pristine road infrastructure is universally recognized, it's equally undeniable that roadways are persistently subjected to degradation [2]. Factors such as climatic extremes, vehicular stress, and natural wear-and-tear all contribute to the deterioration of road surfaces [3]. The consequent damages, ranging from innocuous surface irregularities to perilous potholes, pose significant safety risks to motorists, exacerbate vehicular wear, and escalate maintenance costs. Hence, timely and accurate damage detection is a sine qua non for effective road maintenance and ensuring commuter safety.

Historically, the task of road surface damage detection was primarily relegated to manual inspections. Field engineers and surveyors would periodically inspect stretches of road, logging visible damages for subsequent repair. However, such methods are inherently fraught with shortcomings. Human inspections are not only labor-intensive and time-consuming but are also marked by subjective biases and are often limited by the perceptual constraints of the human eye. Furthermore, large-scale road networks make manual monitoring a logistical challenge, often leading to significant delays between damage occurrence and its eventual rectification [4].

Emerging from this backdrop, technological solutions began to surface, attempting to alleviate the limitations of manual inspection. Early endeavors in this direction exploited image processing techniques to detect road anomalies [5]. While promising, these rudimentary techniques often grappled with issues of low accuracy, particularly in diverse environmental and lighting conditions [6]. More advanced techniques leveraging pattern recognition and machine learning offered an uptick in detection capabilities but remained hamstrung by their inability to perform adequately in real-time scenarios and their frequent misclassifications in complex road environments [7].

The recent upswing in the adoption of deep learning models across diverse domains signaled a transformative potential for road damage detection. Deep learning, a subset of machine learning, empowers models to learn and make decisions from vast amounts of data, often surpassing human-level performance in specific tasks [8]. In the context of road damage detection, Convolutional Neural Networks (CNNs) have emerged as a favored tool due to their adeptness in handling image data [9]. However, while CNNs are proficient in classification tasks, the intricate nature of road damage detection demands a more nuanced approach, one capable of instance segmentation—a task that goes beyond mere classification and seeks to delineate and identify specific objects within images.

This is where the Mask R-CNN model [10] enters the fray. An evolution of the established R-CNN [11] and Fast R-CNN [12] architectures, Mask R-CNN has proven its mettle in instance segmentation tasks across various domains. Its unique architecture, which seamlessly integrates the strengths of both its predecessors, enables precise object localization and pixel-wise mask prediction. Such capabilities render it an intriguing prospect for the intricacies of road surface damage detection.

This research paper aims to exploit the prowess of the Mask R-CNN model in developing a comprehensive framework for real-time road surface damage detection. Drawing upon a meticulously curated dataset encompassing a myriad of road types and conditions, the study seeks to optimize and fine-tune the Mask R-CNN model for this specialized task. Moreover, this investigation delves deep into the challenges inherent in road damage detection, such as variable lighting, shadow effects, wet surfaces, and other environmental nuances. By addressing these complexities, the paper aims to elevate the discourse on automated road damage detection and present a robust, scalable, and efficient solution.

In doing so, this paper positions itself at the intersection of advanced deep learning paradigms and pressing infrastructural challenges. It aspires not just to contribute to academic discourse but also to catalyze tangible shifts in how road maintenance authorities across the globe approach the monumental task of road upkeep and safety assurance. By marrying the Mask R-CNN model's capabilities with the real-world demands of road damage detection, this study embarks on a journey to redefine the standards of road infrastructure management in the age of artificial intelligence.

II. Related Works

The journey of automating road surface damage detection has been a progressive one, punctuated by incremental innovations and paradigm shifts. As this research navigates the waters of the Mask R-CNN model for real-time road surface damage detection, it is imperative to contextualize its approach within the broader framework of previous efforts in this domain. This section endeavors to provide a comprehensive review of related works, elucidating the trajectory of technological advances that have shaped the discourse on automated road damage detection.

A. Traditional Image Processing Techniques

The inception of automated methodologies for road surface damage detection is deeply rooted in traditional image processing techniques. In the nascent stages, simple, yet effective algorithms such as edge detection, thresholding, and morphological operations were employed to discern road anomalies, primarily cracks and potholes. Pioneering research, exemplified by the work of [13], and made strides in this domain by harnessing wavelet transforms for enhanced crack detection. While these early techniques represented a significant leap from manual inspection, they were not without their limitations. Particularly, their susceptibility to variable environmental conditions, such as fluctuating lighting and shadows, frequently resulted in a high rate of false positives. Consequently, despite their foundational contributions, it became evident that more sophisticated approaches were needed to achieve the precision and reliability demanded by real-world applications in road maintenance.

B. Machine Learning and Pattern Recognition

Transitioning from the foundational image processing methodologies, the domain witnessed a paradigm shift with the advent of machine learning and pattern recognition techniques. Here, the emphasis transitioned from raw image manipulation to extracting discernible features, which could then be classified using algorithms. A seminal contribution in this realm was made by [14], who adeptly combined texture-based feature extraction with Support Vector Machines (SVM) to pinpoint road cracks. This strategy elevated the accuracy of detection substantially. However, it also introduced the intricacy of manual feature engineering, a labor-intensive endeavor with potential for inconsistencies. Despite the undeniable advancement in damage detection these methods brought about, the challenges they posed emphasized the need for more automated and adaptive solutions, paving the way for the exploration of deep learning techniques in subsequent research.

C. Deep Learning and CNNs

The renaissance of neural networks, especially Convolutional Neural Networks (CNNs), ushered in a new era for road damage detection. The beauty of CNNs lies in their ability to automatically learn features from raw image data without explicit manual feature engineering. Significant contributions in this realm include the work of [15], who developed a road damage detection and classification system based on deep CNNs. Their model was not only adept at identifying damages but also categorizing them into types like cracks, potholes, and patches. However, while CNNs were proficient in classifying damaged regions, delineating the exact boundaries of these damages remained a challenge.

D. R-CNN and its Evolution

The introduction of Region-based Convolutional Neural Networks (R-CNN) signaled a quantum leap in object detection tasks. R-CNN and its evolutionary offshoots, Fast R-CNN and Faster R-CNN, integrated region proposal networks with CNNs, allowing precise object localization within images [16-18]. In the context of road damage detection, this meant an enhanced ability to identify and demarcate specific damaged regions within a broader road image. The works of [19] stand testament to the efficacy of Faster R-CNN in detecting and segmenting road damages.

E. Instance Segmentation with Mask R-CNN

Delving deeper into the world of object detection, the Mask R-CNN model emerged as a revolutionary tool, bringing the nuance of instance segmentation to the fore. Building upon the foundation laid by its predecessors, the Faster R-CNN, the Mask R-CNN transcended mere object localization, offering pixel-wise mask prediction for each identified entity within an image [20]. This level of granularity made it an optimal candidate for tasks requiring meticulous delineation, such as road damage detection. Early explorations into the model's applicability, highlighted by studies like those of [21], exhibited promising outcomes. The ability of the Mask R-CNN to pinpoint and define road surface anomalies with precision underscored its potential to set a new benchmark in the domain, promising a convergence of accuracy and granularity hitherto unseen in earlier methodologies.

F. Real-Time Detection Challenges

While the evolution of detection techniques marked notable advancements, the exigencies of real-time processing remained a pivotal concern. The operational demands of road maintenance necessitate not just accuracy, but also timeliness in damage detection. Architectures like YOLO (You Only Look Once) and SSD (Single Shot MultiBox Detector), as elucidated by researchers such as [22-23], respectively, heralded solutions emphasizing real-time object detection. Though not tailored explicitly for road anomalies, the underlying principles of these frameworks provide invaluable insights. They spotlight the intricate balance and potential trade-offs between detection speed and accuracy. Such considerations are paramount when envisioning a model that operates in dynamic real-world settings, reinforcing the need for an optimal blend of precision and promptness in any prospective road damage detection system.

G. Adaptive Learning and Transfer Learning

With vast and diverse road networks, training models from scratch becomes computationally expensive. The concept of transfer learning, where models pre-trained on large datasets are fine-tuned for specific tasks, gained traction. The author in [24] explored transfer learning for road damage detection, leveraging models initially trained on datasets like ImageNet and adapting them to the specific nuances of road images.

In synthesizing the above, one discerns a clear trajectory: from basic image processing to the intricacies of deep learning, and from the broad strokes of object detection to the finesse of instance segmentation. This research positions itself at this evolving frontier, seeking to harness the potential of Mask R-CNN, not just in terms of detection accuracy but also in meeting the demands of real-time processing.

III. Materials and Methods

A comprehensive review of pertinent literature underscores the unparalleled efficacy of deep convolutional neural networks (DCNNs) in current scholarly investigations. A pivotal initial step involves segmenting roadway imagery to demarcate relevant classes, crucial for defect identification. Presently, CNN architectures, such as the SegNet [25] and U-Net [26], are gaining traction for their effectiveness in this domain. The challenge arises from the subtle grayscale variations in road imagery and the minimal contrast between the intended subject and its backdrop, compounded by incidental noise and unrelated elements. To navigate these challenges, a fully convolutional neural network (FCNN) employing an "encoder-decoder" configuration is utilized, yielding a binary output image [27]. The FCNN bifurcates into a convolutional segment—transforming the primary image into a feature-rich representation—and a segment producing the segmented output from these features. This architecture encompasses a series of convolutional strata, augmented by filters and subsequent sub-discretization tiers. By integrating upsampling with convolutional stages, the architecture reconstructs the initial image dimensions, subsequently crafting a likelihood matrix.

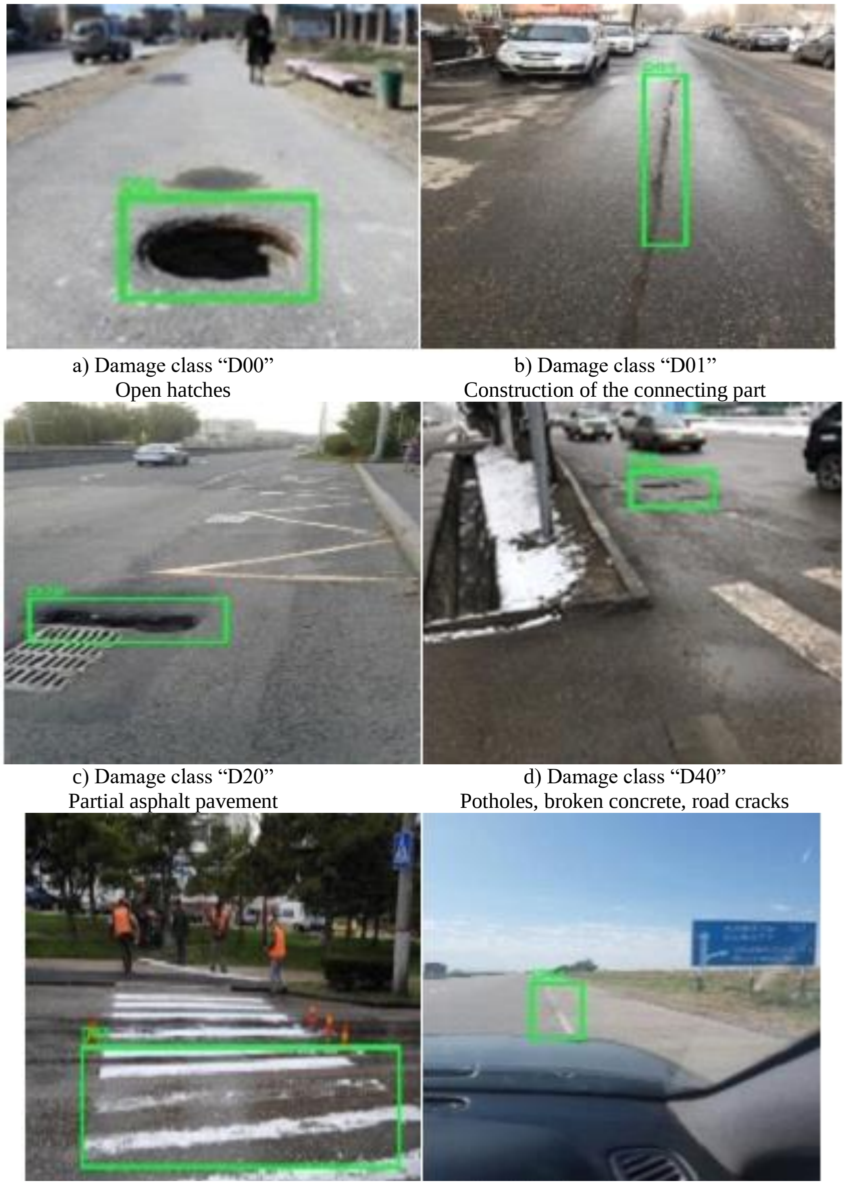
The CrackForest dataset comprises 117 snapshots, partitioned into training, testing, and validation subsets. For each image, 64x64 segments are extracted arbitrarily from both training and test sets. Image quality amplifies with gamma correction, enhancing neural network performance. A 95:5 ratio, emphasizing defects constituting at least 5% of the image, is deemed optimal. With 15,200 training fragments juxtaposed against 3,968 test fragments, the balance is deemed propitious for the deep learning process. The network's evaluation employs intersection over union metrics, complemented by binary similarity metrics. Weight initialization within FCNN layers leverages the Glorot technique, normalizing each layer's input distributions, thus mitigating internal covariance shifts. Optimization ensues via the Adam optimizer. Research concludes that an optimal 25-epoch training duration—split between an initial 5 epochs and a subsequent 20—is effective. Execution of the FCNN blueprint leverages both Keras and TensorFlow platforms. Upon training completion, the artificial neural network undergoes rigorous testing and validation using sample data.

In this study, an enhanced methodology trained existing Mask R-CNN models via TensorFlow's Object Detection API, aiming to augment road defect detection efficiency. These refined models subsequently underwent rigorous evaluations utilizing meticulously curated annotation datasets.

H. Data Collection and Preparation

Traditionally, road surface damage detection relied on aerial images or imagery sourced from vehicle-mounted cameras. Aerial imaging poses practical challenges due to the intricacies involved in capturing such images, restricting its widespread application. Conversely, using imagery derived from vehicle-mounted cameras offers more pragmatic utility, considering the ease of data acquisition. This positions commonly available devices, like smartphones, as potential tools for damage detection, whether the processing occurs in situ or is offloaded to a remote server. Consequently, we developed a unique dataset encompassing six distinct categories of road damage, with each image meticulously annotated by hand.

Fig. 1 presents a visual guide to the diverse damage types, denoted by specific class names such as D20. The subsequent illustrative table segregates these damages into six primary categories, distinguishing between cracks and other deformities. Crack-based damages further bifurcate into linear and alligator cracks, while other categories span potholes, ruts, and anomalies like faded lane markings. Notably, the breadth of damage categories explored in our study outstrips the limited scopes of prior works. For instance, the approach proposed by [28] merely detects potholes under the D40 label, while Jana et al. [29] differentiates damages strictly as longitudinal or transverse. Further, preceding deep learning studies [30-33] primarily focus on identifying the mere presence or absence of damage.



a) Damage class “D00”
Open hatches

b) Damage class “D01”
Construction of the connecting part

c) Damage class “D20”
Partial asphalt pavement

d) Damage class “D40”
Potholes, broken concrete, road cracks

e) Damage class “D43”
Blurring a road crossing

f) Damage class “D44”
Blurring the dividing lines

Fig. 1. Road damage photos and classes for a model training.

I. Annotation and Classification for Enhanced Damage Detection

To facilitate a refined categorization, our annotation data delineates 12 distinct classifications of road damage and associated features captured in the photographs. The Microsoft Visual Object Tagging Tool (VoTT) was instrumental in annotating these color images. Within each image, specifically its lower two-thirds, every discernible feature within our predefined classes was segmented and appropriately labeled. Table I elucidates the compiled annotation data.

Among these classifications, “Scratches on Markings” emerged as the most prevalent, boasting 3,360 segments. This was closely followed by "Linear Cracks" at 3,080 segments. On the rarer end, "Grid Cracks in Patchings" registered the least at 252 segments, succeeded by “Stains”, “Manholes”, and “Potholes”. For analytical rigor, the data segments were stratified into training, validation, and testing datasets at a proportion of 0.6:0.2:0.2, respectively.

In our comprehensive research, we established a refined taxonomy of annotation data that encompasses 12 unique classifications pertinent to road damage and its corresponding features as depicted in the photographic evidence. The intricacies of the annotation process were adeptly managed using the Microsoft Visual Object Tagging Tool (VoTT), which proved pivotal for effective categorization within the color images. A keen focus was directed towards the inferior two-thirds of each image. Within this portion, every feature that aligned with our pre-established categories was diligently segmented and given an appropriate label. For a detailed scholarly overview, readers are directed to Table I, which presents a thorough synthesis of the amassed annotation data.

TABLE I. ROAD IMAGES ANNOTATION DATA

Class ID	Classes	Training	Validation	Testing	Total
1	Linear crack	3080	660	660	4400
2	Grid crack	658	141	141	940
3	Pavement joins	854	183	183	1220
4	Patchings	448	96	96	640
5	Fillings	1344	288	288	1920
6	Pot-holes	406	87	87	580
7	Manholes	336	72	72	480
8	Stains	266	57	57	380
9	Shadow	1190	255	255	1700
10	Pavement markings	1414	303	303	2020
11	Scratches on markings	3360	720	720	4800
12	Grid crack in patchings	252	54	54	360
0	Total	13608	2916	2916	19440

Among these classifications, “Scratches on Markings” emerged as the most prevalent, boasting 3,360 segments. This was closely followed by "Linear Cracks" at 3,080 segments. On the rarer end, "Grid Cracks in Patchings" registered the least at 252 segments, succeeded by “Stains”, “Manholes”, and “Potholes”. For analytical rigor, the data segments were stratified into training, validation, and testing datasets at a proportion of 0.6:0.2:0.2, respectively.

IV. PROPOSED NETWORK

In pursuit of an integrated solution for crack identification and their granular pixel-wise delineation, the contemporary Mask R-CNN convolutional network architecture was chosen. Delving into its foundation and operational mechanics, one finds that the Mask R-CNN is rooted in a lineage of convolutional neural networks designed for localized region processing. This lineage encompasses the Region-based Convolutional Neural Network (R-CNN), its subsequent iterations in Fast R-CNN, and the even more refined Faster R-CNN.

Fig. 2 portrays our adoption of the Mask R-CNN architecture tailored for road surface damage identification. At its core, the Mask R-CNN framework is intrinsically intricate in its block configuration. The initial phase involves the input image being processed through the network, highlighting a feature map. Common feature extractors employed for this purpose include VGG-16, the 50-layer Residual Neural Network (ResNet50), and the more extensive 101-layer Residual Neural Network (ResNet101), with layers focused on classification being omitted. An evolutionary distinction of this architecture, setting it apart from earlier iterations, is the incorporation of the Feature Pyramid Network (FPN) methodology. This technique is pivotal in harvesting feature maps across varied scales. Within this paradigm, consecutive layers of the network, characterized by descending dimensions, are perceived as a stratified "pyramid", where lower tier maps are high-resolution, and the apex tiers possess enhanced semantic abstraction.

Post this feature map extraction, the Region Proposals Network (RPN) segment takes center stage. Its primary objective is to pinpoint hypothesized regions within the image that potentially harbor objects. This is achieved by sliding a 3x3 neural network window over the feature map, with the output anchored on predefined 'k anchors' – essentially frameworks with specified dimensions and orientations. For every such anchor, the RPN forecasts the object's presence and, if detected, fine-tunes the coordinates of the object's bounding box. This stage's ultimate goal revolves around spotlighting regions brimming with potential object presence. Consecutively, overlapping regions are eliminated, courtesy of the non-maximum suppression operation.

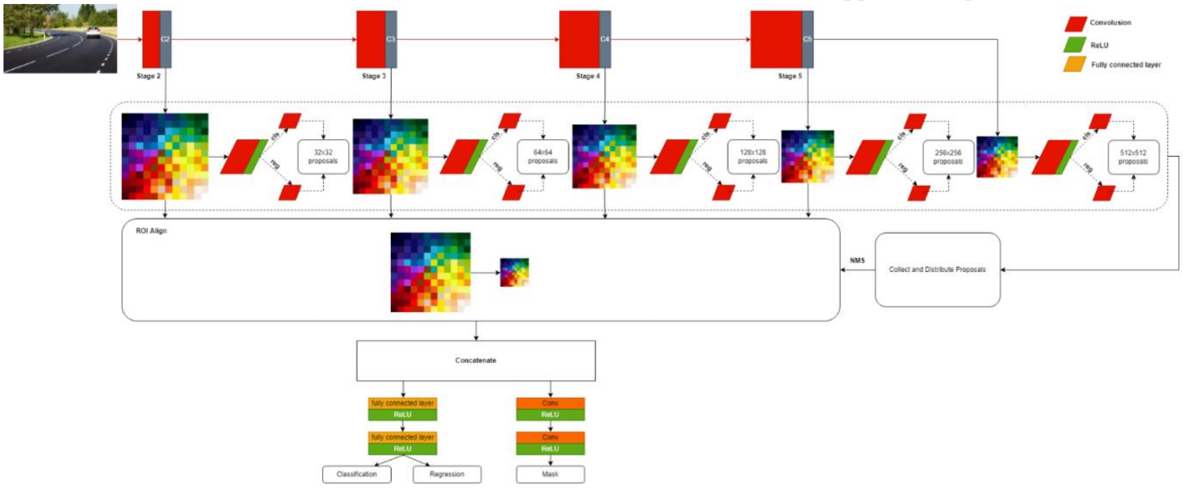


Fig. 2. Proposed mask R-CNN model for road surface damage detection.

In the subsequent phase, the Region of Interest (ROI) Align mechanism comes into play, selecting values pertinent to these regions from the feature maps and standardizing them to a uniform size. These harmonized values then undergo final processes including classification, adjustment of bounding box coordinates, and mask prediction. Notably, the emergent mask, despite its considerably diminished size, retains real values. Once the mask is scaled congruent to the object's dimensions, the precision achieved is commendable.

V. Evaluation

To ascertain the efficacy of the suggested model, it's evaluated against key metrics, specifically the mean average precision (MaP) and the average recall (AR), both scrutinized at varying thresholds of intersection over union (IoU). In scenarios involving classification paired with object localization and detection, the ratio derived from the areas of the bounding boxes frequently serves as a determinant metric, reflecting the accuracy of the bounding box placement.

Embedded within the Mask RCNN is a region proposal network layer, adept at executing parallel inferences concerning class categorization, segmentation, and mask territories, leading to a resultant of six distinctive loss metrics. Complementing these inherent model-specific metrics, both average precisions and average recalls, benchmarked at an IoU value of 0.5, are employed across all twelve delineated road object categories, as referenced in [34].

$$IoU = \frac{S(A \cap B)}{S(A \cup B)} \tag{1}$$

Given A as the forecasted bounding box and B as the reference bounding box, the Intersection over Union (IoU) serves as a metric. The IoU value stands at zero when the bounding boxes do not intersect, and reaches its zenith of one when the bounding boxes perfectly coincide.

A pivotal aim of evaluation is maximizing the detection of instances within a given population using a screening method. It's imperative that false negatives are curtailed, even if it necessitates an uptick in false positives. This emphasis necessitates the careful consideration of three fundamental metrics: the true positive rate (TPR), false positive rate (FPR), and overall accuracy (ACC). Within the realm of medical terminologies, TPR often finds its synonym in sensitivity (SEN) and is represented as seen in equation (2), as documented in [35].

$$TPR = SEN = \frac{TP}{P} \tag{2}$$

Let TP represent the count of true positives, while P signifies the total positive instances in the dataset.

The quantification of the subsequent metric, the false positive rate, is articulated in equation (3), as delineated in [36]:

$$FPR = \frac{FP}{N} \tag{3}$$

Where N denotes the aggregate count of negative cases in the population and FP symbolizes the number of false positives. Furthermore, the true negative instances are also represented by N. However, a more intuitive understanding of this metric is the fraction of true negatives out of the actual negative cases. In medical terminology, this metric is often referred to as specificity (SPEC), articulated as equation (4), as cited in [37]:

$$TNR = SPEC = \frac{TN}{N} = 1 - FPR \tag{4}$$

Ultimately, the metric of accuracy encapsulates the equilibrium between true positive and true negative outcomes. This metric becomes particularly insightful when there exists an imbalance between positive and negative instances within the dataset. This is quantitatively represented in equation (5), as referenced in [38]:

$$ACC = \frac{TP + TN}{P + N} \tag{5}$$

VI. Experimental Results

Within this segment, the experimental findings are bifurcated into two distinct subsections. The initial subsection elucidates the results pertaining to road damage detection, followed by an exposition on road damage segmentation outcomes. The subsequent section delves into the real-time performance of the proposed model, accompanied by visual demonstrations. This encompasses both original imagery and annotated representations of road conditions. In the third subsection, a comprehensive assessment of the model is presented, detailing evaluative metrics such as precision, recall, and F-score for the respective categories of road surface impairments.

A. *Road Damage Detection Results*

Leveraging the intricacies of the Mask R-CNN architectural framework, we developed a nuanced system tailored for road damage detection. This state-of-the-art approach is adept at swiftly and accurately discerning multiple forms of roadway degradation, encompassing anomalies like cracks and spalling, as evidenced in the images procured using digital photographic equipment. To facilitate an insightful understanding and comparison of the system's performance, Table II meticulously catalogs the results of the damage detection endeavor. This tabulation emphasizes evaluative metrics, notably precision, recall, and the F1-score, underscoring the robustness and precision of the devised methodology.

B. *Road Damage Detection Results*

In the process of isolating the segment of the image associated with the roadway, pixels within the road mask are accentuated. Subsequently, an 8-connected region search algorithm is employed on the resultant binary mask. The region boasting the highest pixel count is subsequently identified as the coverage mask, as depicted in a gray shade in Fig. 3.

TABLE II. EVALUATION OF THE PROPOSED METHOD BY CLASSES

Model	Precision	Recall	F1-score
Proposed model	**0.9214**	**0.9876**	**0.9571**
Fully convolutional encoder–decoder network [39]	0.9130	0.9410	0.9270
Deep learning-based semantic segmentation [40]	0.8340	0.6855	0.7524
UNet-based concrete crack detection CrackUnet19 [41]	0.9145	0.8867	0.9004
Two-step light gradient boosting machine [42]	0.6801	0.7578	0.6950
Semantic segmentation using deep learning [43]	0.4044	0.7847	0.4994
Automated vision-based detection [44]	0.9236	0.8928	0.9079

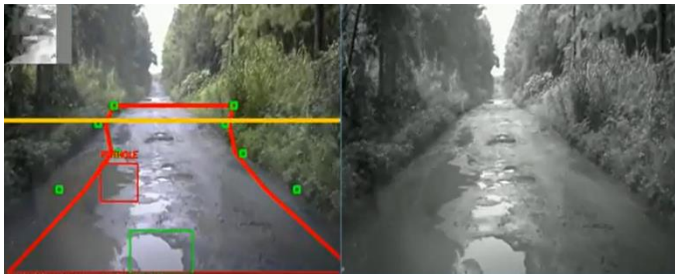


Fig. 3. Marked up road images.

To assess the proficiency of the devised methodology for defect detection, a curated dataset comprising 50 authentic images showcasing road cracks was meticulously assembled. Fig. 4 juxtaposes the outcomes of manual crack delineation against the segmentation outcomes achieved through the proposed neural network's granular pixel-wise selection on an actual image.

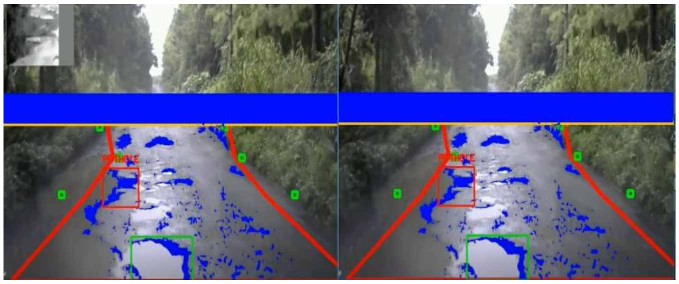


Fig. 4. Marked up pixel-wise selection of road images.

Fig. 5 presents the outcomes of model evaluation over 100 epochs. In Fig. 5, the accuracy and validation accuracy of the advanced model are delineated. It can be inferred from the data that our model achieves an approximate accuracy of 90% within 60 epochs, indicative of its robustness and applicability in real-world scenarios.

Fig. 6 depicts the training and validation loss associated with the model. The observed minimal loss suggests that the model is poised to commit minimal errors in practical applications.

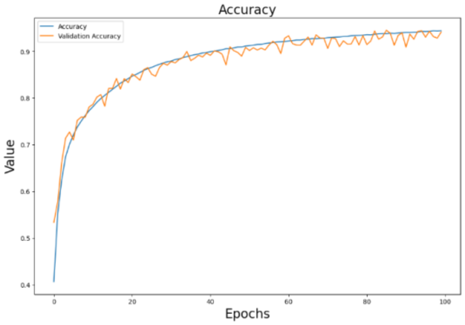


Fig. 5. Accuracy in road damage detection.

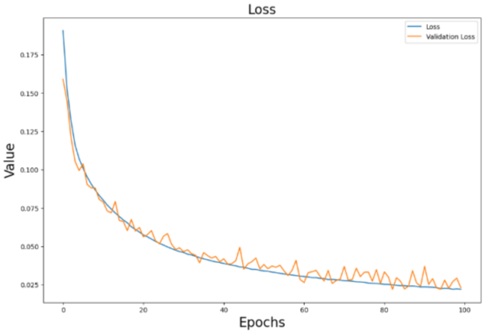


Fig. 6. Loss in road damage detection.

Various strategies employing deep learning paradigms aim to enhance road safety. Contemporary research offers innovative solutions to this issue [45]. For instance, [46] introduced a Vehicle Re-Identification technique to address challenges stemming from significant intra-class variances due to changing vehicle viewpoints during motion and pronounced inter-class resemblances due to analogous appearances. Our model is tailored to identify road surface imperfections using smartphone cameras or any equipment capable of capturing real-time road footage. Based on the results from the conducted experiments, it can be posited that deep learning techniques hold promise in addressing road safety and security challenges.

Table III presents the metrics associated with the model concerning bounding boxes and segmentation masks. For bounding boxes, the metrics for mAP at various IoU thresholds (IoU=.50:.05:.95), mAP (IoU=.50), and mAP (IoU=.75) register as 0.2432, 0.4382, and 0.2482, respectively. In contrast, these metrics for segmentation masks are discerned to be 0.1600, 0.3257, and 0.1279, marking a noticeable decline. The Precision mAP (small) for minuscule objects manifests as markedly lower values, being 0.0365 and 0.0133 for bounding boxes and segmentation masks, respectively, especially when juxtaposed against the Precision mAP for larger and medium-sized entities. The Average Recall metrics for small, medium, and large entities on bounding boxes are quantified as 0.1166, 0.3132, and 0.4717 respectively, whereas the corresponding values for segmentation masks are 0.1021, 0.2528, and 0.2732. Pertaining to our designated damage categories such as linear cracks (denoted as Crack1), grid cracks (labelled as Crack2), potholes, scratches on road markings, and grid cracks in surface repairs, the detection precision metrics at an IoU threshold of .50 are 0.4085, 0.4958, 0.5714, 0.5934, and 0.4000, respectively.

TABLE III. EVALUATION OF THE PROPOSED METHOD BY CLASSES

Classes	**Precision @ 0.5 IoU (Bounding box)**	**Recall @ 0.5 IoU (Bounding box)**	**Recall @ 0.5 IoU (Segmentation)**	**Recall @ 0.5 IoU (Segmentation)**
Linear crack	0.5383	0.3847	0.3583	0.2639
Grid crack	0.6256	0.7140	0.5920	0.6744
Pavement joins	0.4900	0.5179	0.2498	0.2531
Patchings	0.7644	0.5584	0.8161	0.5843
Fillings	0.6071	0.4667	0.3040	0.2528
Pot-holes	0.7012	0.4155	0.7012	0.4155
Manholes	0.9596	0.8798	0.9596	0.8798
Stains	0.1798	0.1484	0.1191	0.1282
Shadow	0.5273	0.4317	0.3285	0.2713
Pavement markings	0.7522	0.7460	0.5065	0.5002
Scratches on markings	0.7232	0.7531	0.4863	0.4944
Grid crack in patchings	0.5298	0.2474	0.7298	0.3063

VII. CONCLUSION

This research delved deeply into the realm of road surface damage detection, harnessing the potential of the Mask R-CNN architecture. The imperative need to develop robust, accurate, and real-time systems for detecting and classifying road damages stems from the crucial role such systems play in ensuring roadway safety and aiding in timely maintenance. A cornerstone of infrastructure management, road health significantly impacts both economic metrics and public safety.

The Mask R-CNN model showcased its prowess in detecting various types of surface damages with commendable precision. Emphasis was placed on understanding its structural nuances and ensuring optimal parameter selection to refine the resultant models. Features like the Region Proposals Network and the integration of the Feature Pyramid Network brought depth and versatility to the proposed method, allowing it to contend with complex road scenarios.

Key metrics used in assessing the model, including mAP and Average Recall across varying IoU thresholds, offered insightful perspectives into the model's performance. The observed results were heartening, with the model showcasing proficiency, especially in differentiating between minor and significant road damage categories.

Comparative analyses with extant literature reinforced the efficacy of the proposed approach, especially considering the challenges posed by real-time, on-ground situations. The model's capacity to work with images and footage from commonplace devices, such as smartphones, stands testament to its applicability in real-world scenarios, democratizing road damage detection to a broader user base.

In summation, while the world of deep learning and neural networks continues to evolve, the application of these technologies in solving pertinent, real-world challenges, as showcased in this study, remains paramount. The presented work not only contributes a robust solution to road surface damage detection but also lays down a pathway for further refinement and innovation in the domain. As future directions, the integration of more advanced architectures and real-time response mechanisms can further elevate the impact and utility of such systems in global infrastructure management.

REFERENCES

[1] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler et al., “The cityscapes dataset for semantic urban scene understanding,” In Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, Nevada, The US, pp. 3213-3223, 2016.

[2] O. Zendel, K. Honauer, M. Murschitz, D. Steininger and G. Dominguez, “Wilddash-creating hazard-aware benchmarks,” In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, pp. 402-416, 2018.

[3] J. Zhang, Y. Sun, H. Liao, J. Zhu and Y. Zhang, “Automatic Parotid Gland Segmentation in MVCT Using Deep Convolutional Neural Networks,” ACM Transactions on Computing for Healthcare, vol. 3, no. 2, pp. 1-15, 2021.

[4] Altayeva, A. B., Omarov, B. S., Aitmagambetov, A. Z., Kendzhaeva, B. B., & Burkitbayeva, M. A. (2014). Modeling and exploring base station characteristics of LTE mobile networks. Life Science Journal, 11(6), 227-233.

[5] K. Gopalakrishnan, S. Khaitan, A. Choudhary and A. Agrawal, “Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection,” Construction and building materials, vol. 157, no. 1, pp. 322-330, 2017.

[6] Altayeva, A., Omarov, B., Suleimenov, Z., & Im Cho, Y. (2017, June). Application of multi-agent control systems in energy-efficient intelligent building. In 2017 Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS) (pp. 1-5). IEEE.

[7] S. Wu, J. Fang, X. Zheng and X. Li, “Sample and structure-guided network for road crack detection,” IEEE Access, vol. 7, no. 1, pp. 130032-130043, 2019.

[8] M. Maniat, C. Camp and A. Kashani, “Deep learning-based visual crack detection using Google Street View images,” Neural Computing and Applications, vol. 33, no. 21, pp. 14565-14582, 2021.

[9] D. Dewangan and S. Sahu, “RCNet: road classification convolutional neural networks for intelligent vehicle system,” Intelligent Service Robotics, vol. 14, no. 2, pp. 199-214, 2021.

[10] M. Masud, M. Hossain, H. Alhumyani, S. Alshamrani, O. Cheikhrouhou et al., “Pre-trained convolutional neural networks for breast cancer detection using ultrasound images,” ACM Transactions on Internet Technology, vol. 21, no. 4, pp. 1-17, 2021.

[11] S. Bang, S. Park, H. Kim, Y. Yoon and H. Kim, “A deep residual network with transfer learning for pixel-level road crack detection,” Network, vol. 93, no. 84, pp. 89-03, 2018.

[12] Y. Chen, H. Wang, W. Li, C. Sakaridis, D. Dai et al., "Scale-aware domain adaptive faster r-cnn," International Journal of Computer Vision, vol. 129, no. 7, pp. 2223-2243, 2021.

[13] D. Quang and S. Bae. "A hybrid deep convolutional neural network approach for predicting the traffic congestion index," Promet-Traffic & Transportation, vol. 33, no. 3, pp. 373-385, 2021.

[14] N. Safaei, O. Smadi, B. Safaei and A. Masoud, “Efficient road crack detection based on an adaptive pixel-level segmentation algorithm,” Transportation Research Record, vol. 2675, no. 9, pp. 370-381, 2021.

[15] S. Bang, S. Park, S., Kim and H. Kim, “Encoder‐decoder network for pixel - level road crack detection in black - box images,” Computer - Aided Civil and Infrastructure Engineering, vol. 34, no. 8, pp. 713-727, 2019.

[16] V. Tran, T. Tran, H. Lee, K. Kim, J. Baek et al., “One stage detector (RetinaNet)-based crack detection for asphalt pavements considering pavement distresses and surface objects,” Journal of Civil Structural Health Monitoring, vol. 11, no. 1, pp. 205-222, 2021.

[17] Z. Lingxin, S. Junkai and Z. Baijie, “A review of the research and application of deep learning-based computer vision in structural damage detection,” Earthquake Engineering and Engineering Vibration, vol. 21, no. 1, pp. 1-21, 2022.

[18] K. Gopalakrishnan, S. Khaitan, A. Choudhary and A. Agrawal, “Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection,” Construction and building materials, vol. 157, no. 1, pp.322-330, 2017.

[19] S. Patra, A. Middya and S. Roy, “PotSpot: Participatory sensing based monitoring system for pothole detection using deep learning,” Multimedia Tools and Applications, vol. 80, no. 16, pp. 25171-25195, 2021.

[20] T. Rateke and A. Von Wangenheim, “Road surface detection and differentiation considering surface damages,” Autonomous Robots, vol. 45, no. 2, pp. 299-312, 2021.

[21] F. Yang, L. Zhang, S. Yu, D. Prokhorov, X. Mei et. al, “Feature pyramid and hierarchical boosting network for pavement crack detection,” IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 4, pp. 1525-1535, 2019.

[22] Q. Zou, Y. Cao, Q. Li, Q. Mao and S. Wang, “CrackTree: Automatic crack detection from pavement images,” Pattern Recognition Letters, vol. 33, no. 3, pp. 227-238, 2012.

[23] Y. Shi, L. Cui Z. Qi, F. Meng and Z. Chen, “Automatic road crack detection using random structured forests,” IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 12, pp. 3434-3445, 2016.

[24] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama and H. Omata, “Road damage detection and classification using deep neural networks with smartphone images,” Computer - Aided Civil and Infrastructure Engineering, vol. 33, no. 12, pp. 1127-1141, 2018.

[25] H. Afify, K. Mohammed and A. Hassanien, “An improved framework for polyp image segmentation based on SegNet architecture,” International Journal of Imaging Systems and Technology, vol. 31, no. 3, pp. 1741-1751, 2021.

[26] B. Omarov, A. Tursynova, O. Postolache, K. Gamry, A. Batyrbekov et al., “Modified UNet Model for Brain Stroke Lesion Segmentation on Computed Tomography Images,” CMC-Computers, Materials & Continua, vol. 71, no. 3, pp. 4701–4717, 2022.

[27] D. Laredo, S. Ma, G. Leylaz, O. Schütze and J. Sun, “Automatic model selection for fully connected neural networks,” International Journal of Dynamics and Control, vol. 8, no. 4, pp. 1063-1079, 2020.

[28] H. Maeda, T. Kashiyama, Y. Sekimoto, T. Seto and H. Omata, “ Generative adversarial network for road damage detection, ” Computer - Aided Civil and Infrastructure Engineering, vol. 36, no. 1, pp. 47-60, 2020.

[29] S. Jana, S. Thangam, A. Kishore, V. Sai Kumar and S. Vandana, “Transfer learning based deep convolutional neural network model for pavement crack detection from images,” International Journal of Nonlinear Analysis and Applications, vol 13, no. 1, pp. 1209-1223, 2022.

[30] B. Kim, N. Yuvaraj, K. Sri Preethaa and R. Arun Pandian, “Surface crack detection using deep learning with shallow CNN architecture for enhanced computation,” Neural Computing and Applications, vol. 33, no. 15, pp. 9289-9305, 2021.

[31] Kulambayev, B., Beissenova, G., Katayev, N., Abduraimova, B., Zhaidakbayeva, L., Sarbassova, A., ... & Shyrakbayev, A. (2022). A Deep Learning-Based Approach for Road Surface Damage Detection. Computers, Materials & Continua, 73(2).

[32] E. Protopapadakis, A. Voulodimos, A. Doulamis, N. Doulamis and T. Stathaki, “Automatic crack detection for tunnel inspection using deep learning and heuristic image post-processing,” Applied intelligence, vol. 49, no. 7, pp. 2793-2806, 2019.

[33] Jayakumar, L., Chitra, R. J., Sivasankari, J., Vidhya, S., Alimzhanova, L., Kazbekova, G., ... & Teressa, D. M. (2022). QoS Analysis for Cloud-Based IoT Data Using Multicriteria-Based Optimization Approach. Computational Intelligence and Neuroscience, 2022.

[34] D. Russo, K. Zorn, A. Clark, H. Zhu and S. Ekins, “Comparing multiple machine learning algorithms and metrics for estrogen receptor binding prediction,” Molecular pharmaceutics, vol. 15, no. 10, pp. 4361-4370, 2018.

[35] V. Thambawita, D. Jha, H. Hammer, H. Johansen, D. Johansen et al., “An extensive study on cross-dataset bias and evaluation metrics interpretation for machine learning applied to gastrointestinal tract abnormality classification,” ACM Transactions on Computing for Healthcare, vol. 1, no. 3, pp. 1-29, 2020.

[36] Sultanovich, O. B., Ergeshovich, S. E., Duisenbekovich, O. E., Balabekovna, K. B., Nagashbek, K. Z., & Nurlakovich, K. A. (2016). National Sports in the Sphere of Physical Culture as a Means of Forming Professional Competence of Future Coach Instructors. Indian Journal of Science and Technology, 9(5), 87605-87605.

[37] Sultan, D., Omarov, B., Kozhamkulova, Z., Kazbekova, G., Alimzhanova, L., Dautbayeva, A., ... & Abdrakhmanov, R. (2023). A Review of Machine Learning Techniques in Cyberbullying Detection. Computers, Materials & Continua, 74(3).

[38] S. Guillon, F. Joncour, P. Barrallon and L. Castanié, “Ground-truth uncertainty-aware metrics for machine learning applications on seismic image interpretation: Application to faults and horizon extraction,” The Leading Edge, vol. 39, no. 10, pp. 734-741, 2020.

[39] M. Islam and J. Kim, “Vision-based autonomous crack detection of concrete structures using a fully convolutional encoder–decoder network,” Sensors, vol. 19, no. 19, pp. 4251, 2019.

[40] T. Yamane and P. Chun, “Crack detection from a concrete surface image based on semantic segmentation using deep learning,” Journal of Advanced Concrete Technology, vol. 18, no. 9, pp. 493-504, 2020.

[41] L. Zhang, J. Shen and B. Zhu, “A research on an improved Unet-based concrete crack detection algorithm,” Structural Health Monitoring, vol. 20, no. 4, pp. 1864-1879, 2021.

[42] P. Chun, S. Izumi and T. Yamane, “Automatic detection method of cracks from concrete surface imagery using two - step light gradient boosting machine, ” Computer - Aided Civil and Infrastructure Engineering, vol. 36, no. 1, pp. 61-72, 2021.

[43] D. Lee, J. Kim and D. Lee, “Robust concrete crack detection using deep learning-based semantic segmentation,” International Journal of Aeronautical and Space Sciences, vol. 20, no. 1, pp. 287-299, 2019.

[44] B. Kim and S. Cho, “Automated vision-based detection of cracks on concrete surfaces using a deep learning technique,” Sensors, vol. 18, no. 10, pp. 3452, 2018.

[45] Omarov, B., Narynov, S., Zhumanov, Z., Gumar, A., & Khassanova, M. (2022). A Skeleton-based Approach for Campus Violence Detection. Computers, Materials & Continua, 72(1).

[46] X. R. Zhang, X. Chen, W. Sun, X. Z. He, “Vehicle Re-Identification Model Based on Optimized DenseNet121 with Joint Loss”, Computers, Materials & Continua, vol. 67, no. 3, pp. 3933-3948, 2021.