

PROBABILISTIC MANAGEMENT OF PAVEMENT DEFECTS WITH IMAGE PROCESSING TECHNIQUES

Felix OBUNGUTA¹, Kakuya MATSUSHIMA², Junichi SUSAKI³

¹Department of Civil Engineering, Graduate School of Engineering, Osaka University, Osaka, Japan

²Disaster Prevention Research Institute, Kyoto University, Kyoto, Japan

³Department of Civil and Earth Resources Engineering, Graduate School of Engineering, Kyoto University, Kyoto, Japan

Article History:

- received 18 January 2023
- accepted 11 October 2023

Abstract. Pavement management has traditionally relied on human-based decisions. In many countries, however, the pavement stock has recently increased, while the number of management experts has declined, posing the challenge of how to efficiently manage the larger stock with fewer resources. Compared to efficient computer-based techniques, human-based methods are more prone to errors that compromise analysis and decisions. This research built a robust probabilistic pavement management model with a safety metric output using inputs from image processing tested against the judgment of experts. The developed model optimized road pavement safety. The study explored image processing techniques considering the trade-off between processing cost and output accuracy, with annotation precision and intersection over union (IoU) set objectively. The empirical applicability of the model is shown for selected roads in Japan.

Keywords: pavement management, image processing, objective annotation and IoU, expert validation.

 Corresponding author. E-mail: obongutafelix@gmail.com

Notations

Variables and functions

A – intervention;
 k – section;
 $\text{cost}_{A \rightarrow k}$ – intervention cost for A on k ;
 d_k^n – defect density of class n for k ;
 \mathbf{d}_k – vector of defect densities for k ;
 S_k – safety metric;
 i – annotation case;
 n_i – number of annotation cases;
 e – expert;
 C_k – cracking ratio for k ;
 D_k – rutting/pothole depth for k ;
 σ_k – roughness for k ;
 MCI_k – MCI for k ;
 MCI_{\min}^k – minimum MCI for k ;
 MCI_{\max}^k – maximum MCI for k ;
 MCI_k^* – MCI obtained for optimum i^* and IoU^* ;
 MCI_k – safety limit MCI;

p – probability of MCI falling within a given range;
 p_0 – probability limit;
 β_q – unknown parameters;
 $\boldsymbol{\beta}$ – vector of unknown parameters;
 $\delta_i, \delta_{k,e}$ – dummy variables.

Abbreviations

ANN – artificial neural network;
 CNN – convolutional neural network;
 GPS – global positioning system;
 IoU – intersection over union;
 LCC – life cycle costs;
 mAP – mean average precision;
 Mask R-CNN – mask region convolutional neural network;
 MCI – Maintenance Control Index;
 PMS – Pavement Management System;
 RDD – Road Damage Dataset;
 RoI – region of interest;
 YOLOv4 – You Only Look Once version 4.

1. Introduction

Pavement monitoring and management faces a number of challenges including a shortage of experts for an increasing pavement stock, leading to the neglect of pavements (Maeda et al., 2018). This pavement neglect, which is mainly due to inadequate resources, may affect road user safety. Amid this challenge, technological developments such as low-cost smartphones and advanced analysis software including image processing techniques can be leveraged by pavement managers to increase the pavement management scope cost-effectively and efficiently. Additionally, human-based monitoring is prone to errors such as misreporting and is susceptible to bias due to its subjectivity, which may result in different inspection results by different experts for the same sections. Also, large inspection machines are costly and create negative externalities such as traffic delays/disruptions due to closed roadways (Mizutani et al., 2020). This research explored the application and adoption of image processing techniques and deep learning to pavement asset management to improve the management scope and estimation accuracy in order to enable more appropriate management decisions.

Simpler image processing methods exclude the cumbersome and costly image annotation step; however, these methods may be challenged by poorer object detection and segmentation. On the other hand, deep learning that involves costly but vital preliminary steps including image annotation and the setting of model inputs such as intersection over union (IoU) has been shown to be more accurate (He et al., 2018). Because image annotation is done manually by an annotator, it is not only cumbersome but also highly subjective, which affects model learning and object detection. The IoU is an important input parameter for object detection tasks because a very high IoU (more strict) leads to poor detection in which important objects such as pavement defects may be undetected, whereas a very low IoU (less strict) may result in detection of unwanted objects including roadside vegetation. To avoid erroneous or meaningless detection, very high or very low IoUs are undesirable. For road pavement defect detection tasks, very precise annotations may not be required because the condition need only fall within a given range for effective management decisions to be made. Therefore, a trade-off may exist between the annotation cost and quality as described and modeled herein. This study built a model in which the IoU and annotation quality are set objectively and validated by experts to aid appropriate object detection for specific purposes and shows an empirical application of the probabilistically determined deep learning output to asset management for selected roads in Japan.

2. Background

2.1. Pavement management

Pavement management decisions may be based on the predicted performance of pavement structures. Infrastructure performance models can be placed into three broad

categories: stochastic (probabilistic), deterministic, and computer techniques (Tsuda et al., 2006; Kobayashi et al., 2010; Tabatabaee & Ziyadi, 2013; Pérez-Acebo et al., 2019; Obunguta & Matsushima, 2020). The Bayesian approach has also been used to improve the prediction of infrastructure performance through updating whenever additional data is available (Kobayashi et al., 2012; Tabatabaee & Ziyadi, 2013). Infrastructure asset management is heavily dependent on infrastructure condition, which requires significant amounts of data. Stochastic and deterministic techniques may require a minimum of two-point data to predict infrastructure system performance; however, cases of incomplete data including one-point data occur due to a lack of human resources and equipment to carry out surveys in a wider geographical area and for a larger infrastructure stock. Lethanh and Adey (2012) applied the improved stochastic hidden Markov model for modeling pavement deterioration in case of incomplete monitoring data. Additional data may be generated through multiple imputations (Rubin, 1976, 1987) and/or computer techniques could be used to process one-point data and output useful information to support management decisions (Maeda et al., 2018; Zou et al., 2022).

In the past, data was collected by engineers through periodic inspection, which is prone to errors such as misreporting, omissions, and/or incorrect data entries, especially as the infrastructure stock increases. Human-based detection and measurement of defects is a highly subjective process susceptible to bias. Furthermore, the collected data is normally sorted manually by a data analyst to eliminate unusable data, a process that may introduce additional errors. The poor data problem is further augmented at the data cleaning stage, where a large amount of data is eliminated, affecting the power of estimates obtained from prediction models and thereby blurring management decisions. Barchard and Pace (2011) statistically showed the devastating effects of data errors, prevalent in human-based systems, on estimated results. A shortage of experts has also resulted in reduced inspection coverage leaving significant amounts of infrastructure neglected (Maeda et al., 2018). Additionally, the use of expensive specialized equipment for damage measurement may not be feasible in certain settings and disrupts traffic flow causing delays due to closed roadways (Mizutani et al., 2020). More accurate and effective computer-based infrastructure management using fewer resources (both human and material) is thus desirable.

Pavement management decisions are made to minimize costs, typically life cycle costs (LCC), for a projected period of operation. Kobayashi et al. (2013) developed a pavement management model that optimized pavement inspection and repair by minimizing LCC taking into account pavement soundness and risk control level. Obunguta and Matsushima (2020) optimized pavement system LCC considering different management policies, i.e., time-dependent and condition-dependent, and explored the effect of preventive maintenance on LCC. Efficient pavement intervention may also be determined by optimizing road

usage and utility (K. Lin & C. Lin, 2011; Liu & Wang, 2016; Mizutani et al., 2020).

Pavement Management Systems (PMS) such as the Highway Development and Management Tool (HDM-4) developed by the World Bank have been applied for project appraisal and analysis of road management and investment alternatives through LCC optimization. The HDM-4 incorporated detailed investigative studies in Kenya, India, Brazil and the Caribbean and has been applied in many countries to improve road travel including traffic congestion and road safety (Kerali, 2000). The Kyoto Model PMS developed based on studies including Tsuda et al. (2006) and Kobayashi et al. (2010, 2013) has been applied in various countries including Japan and Vietnam to enhance road asset management (Thao et al., 2015). These PMS models have traditionally used numerical data collected through visual surveys. As highlighted by Maeda et al. (2018), a shortage of experts has hindered data collection necessitating the improvement of these PMSs to make them less human-dependent and more semi-autonomous.

Past studies including Bosurgi et al. (2022) and Georgopoulos et al. (1995) made attempts in using digital image processing data for pavement distress evaluation. Georgopoulos et al. (1995) developed a method that generates digital imagery used as input for a specialized software to determine the severity of pavement cracking. Bosurgi et al. (2022) used computer vision to process road pavement three-dimensional data to output quantified distress measures, i.e., area, perimeter and depth with the goal of using the measures to optimize road functionality and safety. Efficient digitalized systems have the potential to significantly reduce pavement monitoring and management costs, and traffic disruptions when large equipment is used.

2.2. Image processing techniques

2.2.1. Datasets, simple segmentation, and deep learning

Recent technological advancements have led to the development of smartphones with comparatively lower cost but high quality, which has resulted in the production of abundant smartphone road image data. The images may be stored in datasets such as ImageNet (Deng et al., 2009), Pascal VOC (Everingham et al., 2015), CamVid (Cambridge University, 2021), and Road Damage Dataset 2020 (RDD-2020) (Arya et al., 2020a).

For infrastructure systems such as road networks, images are normally collected by taking photos through the car windshield using a smartphone mounted on the dashboard (Figure 1). Car windshield images are complex because they contain considerable noise (many objects) and are in perspective view. The images may additionally be affected by weather, e.g., lighting and shadows. Plan view images, taken directly above the road surface such as by using a drone, may be simpler but are legally prohibited in many jurisdictions.



Figure 1. Setup of smartphone in car (Arya et al., 2021)

Image processing has been applied to many fields such as forestry to evaluate the impacts of policies addressing deforestation (MathWorks Inc., 2021), transportation infrastructure for road damage detection (Maeda et al., 2018; Arya et al., 2020a, 2020b, 2021), and dermatology to determine the severity of skin cancer (Kinyanjui et al., 2019) and skin lesions (Mirikharaji et al., 2021). Zou et al. (2022) applied deep learning using the You Only Look Once version 4 (YOLOv4) algorithm to detect defects in structures after an earthquake disaster. Maeda et al. (2018) developed a road damage detection system using smartphone images in Japan. Thuyet et al. (2022) built an autonomous road inspection system using deep learning and data obtained utilizing a laser crack measurement system (LCMS) to detect cracks and patches. Other studies such as Goncalves and Givigi (2016) and Hong et al. (2020) developed methods to detect and measure crack defects in civil infrastructure from simple image data containing a few objects.

Object recognition systems can be broadly divided into three groups. First, human-based methods, where an inspector observes and measures defects using traditional measurement equipment; second, microscopic inspection using specialized tools; and third, machine vision, in which defects are identified and quantified automatically by image analysis. Machine vision has proved to be the most efficient and accurate of the three. Within machine vision, artificial neural networks (ANNs), convolutional neural networks (CNNs) and pattern recognition using color models have emerged as the most popular (Goncalves & Givigi, 2016). Color models were an advancement of simple threshold segmentation (e.g., Otsu, 1979). Other segmentation methods have been developed, including graph-based segmentation using the lazy snapping technique and region growing from a seed point (MathWorks Inc., 2021).

In deep learning, algorithms built using region proposals and CNNs (R-CNN) have achieved higher accuracy. The Mask R-CNN algorithm (He et al., 2018), an advancement of Faster R-CNN (Ren et al., 2015), Fast R-CNN (Girshick, 2015) and R-CNN (Girshick et al., 2014), is the current state-of-the-art algorithm in the family of object detection and segmentation algorithms using region proposals. The Mask R-CNN algorithm extended Faster R-CNN by adding a branch for predicting segmentation masks from each

region of interest (RoI) and also replaced the RoIPool layer with the quantization-free RoIAlign layer, which solved the misalignment challenge in earlier algorithms.

Object detection techniques build a bounding box based on the object category, as shown in Figure 2, and therefore do not provide information about defect size and shape. Image detection and segmentation algorithms such as Mask R-CNN provide a pixel-wise mask for an object, which gives more details about its shape and size, and may also be more suitable for segmenting complex images containing overlapping objects and different colors, textures, contrasts, and light intensities. The severity of defects, obtained through quantification, is important for the asset management decision process, especially considering user safety.

The main advantage of simple segmentation methods over deep learning is that the former does not require the costly image annotation step; however, these methods are challenged by inaccurate object detection and segmentation (Otsu, 1979; MathWorks Inc., 2021). On the contrary, despite requiring the costly annotation step, deep learning has proven to be more accurate and efficient in performing object detection and segmentation tasks (He et al., 2018).

2.2.2. Image annotation

Despite being labor-intensive, annotation is a vital preliminary step before training a deep learning model and therefore should be done as efficiently and accurately as possible. For object detection, bounding boxes and object labels are manually added to the images at every instance that an object is identified by the annotator. For object segmentation, a pixel-wise mask and object label are manually added to the images at each instance. Higher quality and more precise annotations may increase accuracy; however, a higher time cost is required to achieve such annotations. Past studies including Greenwald et al. (2022) have attempted to optimize the time cost for annotating images by combining expert, crowd, and computer input while ensuring that the required accuracy levels are met. Xu et al. (2021) applied partial annotation to leverage the advantages of using annotated and unannotated regions in the training process for crowd counting tasks. More informative annotations may involve detailed manual boundary drawings for a feature of interest, whereas less informative approximate annotations (e.g., bounding boxes or simplified polygons) may require simpler drawings. Therefore, a trade-off may exist between the quality and time cost of annotating images (Mirikharaji et al., 2021). For infrastructure performance evaluations, the accuracy requirements for measurements may not be as strict compared to fields such as health because classification of the defect level need only fall within a specified range to generate useful information for effective decision making. Practitioners may decide the needed annotation quality for specific purposes more efficiently based on the accuracy–time cost trade-off.

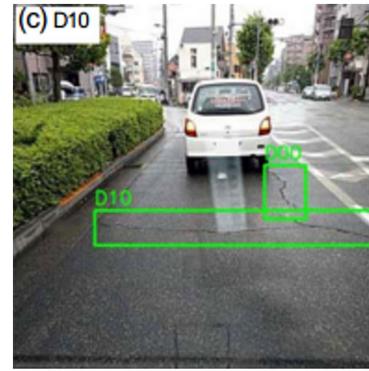


Figure 2. Crack detection (Maeda et al., 2018)

2.3. Problem statement

As discussed earlier, road infrastructure asset management is faced with the challenge of an increasing infrastructure stock with a declining number of management experts, resulting in low coverage and infrastructure neglect, which may adversely affect road user safety. This challenge is further exacerbated by the inaccuracy of human-based inspection and data preparation, which compromises pavement performance results and subsequent management decisions. Road infrastructure asset management should capitalize on gains in the technology industry that has seen the development of low-cost high quality smartphones. Pavement smartphone images from a wider road infrastructure stock can be taken and analyzed using more efficient and accurate image processing techniques compared to human-based methods. For image processing, less accurate simple segmentation methods that do not require costly annotations and the more accurate deep learning process that requires costly annotations are available. A trade-off between accuracy and annotation cost may therefore be evaluated by varying annotation precision requirements from no annotation (for simple segmentation methods) to less detailed annotations and more precise annotations (for deep learning). This trade-off may be used to evaluate the most appropriate image processing methods and requirements for pavement defect detection tasks because very precise defect measurements are unnecessary in pavement asset management since the defects need only fall within a specified range for appropriate intervention prescription. Also noted in the previous sections is the subjectivity of annotation and deep learning model inputs such as IoU, which this research work attempts to set objectively by building a probabilistic pavement asset management model that is validated by expert analysis. Pavement defects are detected and quantified by probabilistically setting IoU and annotation precision followed by effective management and intervention decisions. This paper shows the substitutability of expert analysis with the cost-effective image processing and details how the output from image processing/ deep learning can be applied to pavement asset management to encourage the adoption of more efficient technologies in the asset management practice.

2.4. Research objectives

The main objective of this study was to explore the possibility of arriving at sound pavement management decisions with minimal human dependence. Specifically, the objectives were as follows:

- 1) Carry out an experimental comparison between simple segmentation methods and deep learning.
- 2) Develop a probabilistic pavement management model based on safety and set the annotation precision and IoU objectively including expert validation.
- 3) Empirically show the applicability of the model using image processing output from the processed RDD-2020.

To the best of our knowledge, this is the first study to develop a model that sets the IoU and annotation precision objectively including expert validation and empirically shows the applicability of the probabilistically determined image processing outputs as inputs for asset management decisions. The rest of this article is organized as follows. Section 2 comprehensively explains the probabilistic asset management model; Section 3 presents the empirical model application including deep learning on the RDD-2020; and Section 4 concludes the article and suggests possible future work.

3. Probabilistic pavement asset management model

3.1. Model definition and overview

Consider that a road pavement section k ($k = 1, 2, \dots, K$) has defect density d_k^n estimated from processed image data with n ($n = 1, 2, \dots, N$) indicating the class of the defects, e.g., cracks and potholes. From the estimated d_k^n , an input vector $\mathbf{d}_k = [d_k^1, \dots, d_k^N]$ can be created that generates an output, the safety metric $S_k = f(\mathbf{d}_k)$. The safety metric may be the Maintenance Control Index (MCI) that is commonly used in Japan (JARA, 2013). Each section k can thus have an estimated MCI_k . Based on the severity of the MCI_k , road managers may propose the appropriate intervention A on a section with the following options: do nothing A_0 , crack sealing or patching A_1 , overlay A_2 , and reconstruction A_3 . The estimated defect density d_k^n may vary based on image processing, i.e., set annotation case i and IoU threshold. The defect density may also vary due to other factors including image quality such as the lighting conditions, e.g., shadows; scale, e.g., perspective view; and/or pavement infrastructure properties, e.g., material colors. In this study, the processing methods were emphasized and the goodness of fit of annotation case i and IoU in defect detection was validated using the pavement condition estimates by experts e ($e = 1, 2, \dots, E$). The priority for intervention on a section k is determined based on the magnitude of the safety metric on that particular section in comparison with other sections and the intervention is proposed following an intervention prioritization is

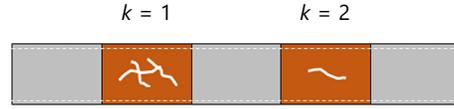


Figure 3. Illustration of two pavement sections

illustrated in Figure 3 with $MCI_{k=1} < MCI_{k=2}$, which implies that the worst section $k = 1$ should receive priority for intervention. The defect densities used to estimate MCI_k are obtained probabilistically by maximizing the probability of detection of defects considering different annotation cases i and IoU thresholds against the expert benchmark. The number of annotation cases is n_i and MCI_k^* is the MCI determined after optimizing i and IoU. The road agency's goal is to maximize the total MCI_k^* for all sections to preserve road user safety within set constraints.

MCI_k is defined as follows (Minami & Suzuki, 2008; Miyamoto & Yoshitake, 2009; JARA, 2013; Yoshida, 2016; Kubo, 2017):

$$MCI_{k,0} = 10 - 1.48C_k^{0.3} - 0.29D_k^{0.7} - 0.47\sigma_k^{0.2}; \quad (1a)$$

$$MCI_{k,1} = 10 - 1.51C_k^{0.3} - 0.30D_k^{0.7}; \quad (1b)$$

$$MCI_{k,2} = 10 - 2.23C_k^{0.3}; \quad (1c)$$

$$MCI_{k,3} = 10 - 0.54D_k^{0.7}; \quad (1d)$$

$$MCI_k = \min(MCI_{k,j}), \quad j(j = 0, \dots, 3), \quad (1e)$$

where C_k is the cracking ratio in percent, D_k is the rutting depth in millimeters, and σ_k is the roughness in millimeters.

3.2. Probabilistic annotation and IoU setting

Consider that the quality of image annotations and the set IoU applied in deep learning can be varied. Assume that the defect density \mathbf{d}_k for a given annotation case i and IoU can be estimated using a computer and can give an output MCI_k . Consider that experts e ($e = 1, 2, \dots, E$) analyze the same K images and grade them using a similar point scale used for computer analysis. Considering the expert analysis as the benchmark, a successful computer match is established if MCI_k falls within $[MCI_{\min}^k, MCI_{\max}^k]$ from the expert analysis; otherwise it is considered a failure (no match). Using multiple logistic regression, the probability $\in [0,1]$ of predicting a binary outcome (match = 1, or no match = 0) can be estimated given the annotation case i and IoU. The explanatory variable i may be considered as categorical and the IoU as continuous within the limits $[0,1]$.

$$\Pr\{MCI_{\min}^k \leq MCI_k \leq MCI_{\max}^k | i, IoU\} = p(i, IoU); \quad (2)$$

$$p(i, IoU) = \frac{\exp(\beta_0 + \beta_1 \delta_1 + \dots + \beta_{n_i-1} \delta_{n_i-1} + \beta_Q IoU)}{1 + \exp(\beta_0 + \beta_1 \delta_1 + \dots + \beta_{n_i-1} \delta_{n_i-1} + \beta_Q IoU)}, \quad (3)$$

where $\boldsymbol{\beta}(\beta_0, \beta_1, \dots, \beta_Q)$ is a vector of unknown parameters to be estimated with q ($q = 0, 1, \dots, Q$) denoting the number of explanatory variables, and δ_i ($i = 1, \dots, n_i - 1$) are dummy

variables for the annotation case. The base annotation case I has all the dummies equal to zero. For three annotation cases, $\delta_1 = \delta_2 = 0$ for the base case I, $\delta_1 = 1$ and $\delta_2 = 0$ for case II, and $\delta_1 = 0$ and $\delta_2 = 1$ for case III. The dummies δ_1 and δ_2 are comparative between cases II and III with the base case, respectively. This definition of dummies is important to avoid indeterminate model coefficients that may occur due to singularities as a result of violating the perfect collinearity property.

Assuming that the probabilities of detection are mutually independent, then the log-likelihood expressing the joint probability density of successful defect detection considering all experts and sections is as follows:

$$\begin{aligned} \ln[\mathbf{L}(\boldsymbol{\beta})] &= \\ \ln \left[\prod_{e=1}^E \prod_{k=1}^K \{p(i, loU)\}^{\delta_{k,e}} \{1-p(i, loU)\}^{(1-\delta_{k,e})} \right] &= \\ \sum_{e=1}^E \sum_{k=1}^K \{ \delta_{k,e} \ln[p(i, loU)] + & \\ (1-\delta_{k,e}) \ln[1-p(i, loU)] \}; & \quad (4) \\ \delta_{k,e} = \begin{cases} 1 & \text{if match} \\ 0 & \text{if no match} \end{cases} & \end{aligned}$$

where $\delta_{k,e}$ is a dummy variable for match or no match.

The unknown parameters can be obtained by maximizing the log-likelihood function:

$$\begin{aligned} \frac{\partial \ln[\mathbf{L}(\boldsymbol{\beta})]}{\partial \beta_q} &= 0; \quad (5) \\ (q = 0, 1, \dots, Q). & \end{aligned}$$

An iterative method such as the Newton–Raphson method could be used to iteratively estimate $\boldsymbol{\beta}$ within a given tolerance level.

For a given annotation case i and loU to be acceptable to effectively detect and quantify defects, the probability $p(i, loU)$ should not be less than a set limit p_0 :

$$p(i, loU) \geq p_0. \quad (6)$$

The annotation case i^* and loU^* that optimizes the probability of the detection of defects is obtained as:

$$\operatorname{argmax}_{i, loU} p(i, loU). \quad (7)$$

The defect densities d_k^{n*} obtained for i^* and loU^* are used in the calculation of MCI_k^* .

3.3. Pavement intervention planning

The choice of intervention A such as patching and overlay on a section may be determined by maximizing the MCI for the entire pavement stock (Obunguta et al., 2022). When action is carried out, it is assumed that the defect density improves and $d_k^{n*} = 0$. The intervention on a given

section is determined based on the MCI following an intervention matrix as shown in Table 1. This type of intervention decision is used by a number of agencies including the Ministry of Land, Infrastructure, Transport and Tourism of Japan (Miyamoto & Yoshitake, 2009; Kubo, 2017). The cutoff level for each defect class intervention can be varied by a road agency based on their standards.

For intervention planning, sections can be prioritized by maximizing the total MCI (safety) subject to constraints. The MCI optimization is defined as:

$$\begin{aligned} \text{Max } & \sum_{k=1}^K MCI_k^*; \quad (8) \\ \text{subject to } & \end{aligned}$$

$$MCI_k^* \geq \overline{MCI}_k; \quad (9a)$$

$$\sum_{k=1}^K \text{cost}_{A \rightarrow k} \leq \text{Budget}, \quad (9b)$$

where \overline{MCI}_k is the safety limit and $\text{cost}_{A \rightarrow k}$ is the intervention cost.

The objective function above includes a serviceability and budget constraint. In case of a budget limit, sections could be prioritized by intervening on those that would result in larger MCI (safety) gains.

In MCI optimization, a Pareto frontier may occur where the MCI_k^* value is the same for two or more sections. In this case, other factors may be considered such as the importance of a pavement section relative to others. If other factors are insignificant, then the prioritization of intervention for the sections at the Pareto frontier may be done randomly.

Detailed pavement intervention planning could be a subject of future research by the authors in which stochastic deterioration estimation considering two-point image data and LCC analysis will be done. The generated pavement deterioration rates could inform better intervention planning through MCI and LCC optimization (Obunguta & Matsushima, 2020). Furthermore, the incorporation of image processing will improve and ease the application of infrastructure asset management models such as HDM-4 and the Kyoto Model by making them more semi-autonomous.

Table 1. Intervention matrix (Miyamoto & Yoshitake, 2009; Kubo, 2017)

MCI_k^*	Intervention			
	A_0	A_1	A_2	A_3
$MCI_k^* = 10$	o	x	x	x
$4.5 \leq MCI_k^* < 10$	x	o	x	x
$2.5 \leq MCI_k^* < 4.5$	x	x	o	x
$MCI_k^* < 2.5$	x	x	x	o

Note: o indicates intervention and x indicates no intervention.

4. Empirical application

4.1. Outline of application

In the empirical application, simple object segmentation methods that do not require costly annotation were explored, and a deep learning model was trained using the Mask R-CNN algorithm in Python 3.9.1 to detect and quantify defects and road features (Rols) in images obtained from the RDD-2020 to estimate defect densities applied in the probabilistic asset management model. To show model practicality, an empirical application was carried out for selected roads in Japan.

4.2. Road image dataset

The RDD-2020 contains images of 600×600 pixels for road surfaces approximately 10 m ahead taken using a camera mounted on a vehicle traveling at an average speed of about 40 km/h (about 10 m/s) and capturing an image every second. The dataset is heterogeneous with more objects and includes images from India, Japan, and the Czech Republic (Figure 4).

4.3. Simple segmentation methods

4.3.1. Overview of methods

There are several segmentation techniques that can be applied to extract features of interest from images. Simple segmentation techniques do not require the costly annotation step. For many images, segmentation needs to be done programmatically as opposed to manually due to

computational cost reasons. This study explored graph-based segmentation using the lazy snapping technique and region growing from a seed point. The segmentation algorithms can be developed in MATLAB and looped through images stored in a specified file directory.

For the lazy snapping technique, the initial background and foreground Rols are user dependent. After the Rols are set, the algorithm programmatically classifies other unallocated image pixels as either background or foreground based on a similarity metric. In Figure 5, consider a 5×5 pixel image with a low pixel (dark) foreground and a high pixel (light) background. The foreground (object) region can be segmented out by specifying the Rol with dimensions [xmin, ymin, width, height] and a background Rol with its own dimensions. The Rols for each group (fore or background) can be as many as necessary. The lazy snapping formula can then be used to group pixels based on similarity.

For region growing from seed point(s), the Rol is iteratively grown by comparing all unallocated neighboring pixels to the Rol based on a similarity measure, as illustrated in Figure 6 where the initial user-dependent seed point S_p with coordinates [x, y] is grown to cover the low pixel object region (Kroon, 2021).

To run either algorithm programmatically, the Rols or the S_p must be pre-set by the user for the entire dataset. When performed programmatically, the algorithms may generate inaccurate results for cases in which the initial Rols and S_p s do not fall in the pixel area of the feature of interest for all images.



Figure 4. Road images from: a – Japan; b – India; c – Czech Republic

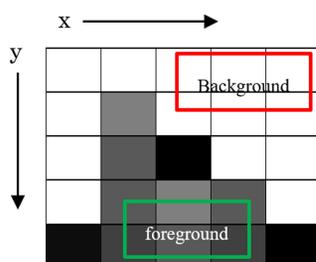


Figure 5. Graph-based segmentation by lazy snapping with a foreground and background Rol

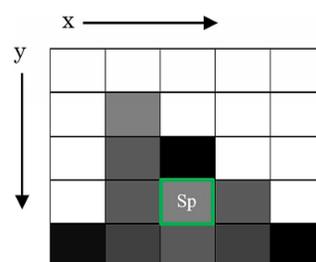


Figure 6. Region growing from a seed point S_p

4.3.2. Simple segmentation experiments

The segmentation experiments shown in Figure 7 and Figure 8 highlighted the fact that both lazy snapping and region growing from a seed point were challenged by the following:

- 1) Region continuity breakage due to lighting conditions, i.e., dark shadows (Figure 8a), bright shiny surfaces, and bright reflections (Figure 7c and Figure 8c) probably from the windscreen due to the camera flash.
- 2) Breakage in segmentation regions due to the color difference between the road markings such as zebra crossings and lane separations, and the pavement surface (Figure 7b, 7c and Figure 8b, 8c).
- 3) Segmentation of unwanted regions (Figure 7d).

The simple segmentation experiments showed that both techniques were challenged by the complexities of segmenting features as a result of lighting and color changes, and unwanted objects were segmented in several cases because segmentation of complex images is challenged by inaccurate initial RoI specification if done programmatically. On the other hand, manual segmentation may be cumbersome; hence, deep learning may offer more accurate segmentation results despite having a higher computational cost (annotation and training) compared to simple segmentation methods.

4.4. Deep learning

4.4.1. Algorithm

Deep learning involves annotating images and then training a model to detect the annotated RoIs. In this study, the deep learning model was trained using the Mask R-CNN

algorithm to detect and build a pixel-wise mask on road features and defects. The main steps of the algorithm are detailed in Table 2 below.

Table 2. Deep learning algorithm

Algorithm: Deep learning and defect quantification
Start
Step 1: Obtain road pavement images
Step 2: Sort images
Step 3: Annotate images in the training and validation set
Step 4: Input annotated images then train and test the deep learning model
Step 5: Defect quantification
Step 6: Output quantified defects; i.e., defect densities
End

4.4.2. Deep learning accuracy

The mean average precision (mAP) is a popular metric in computer vision for evaluating the accuracy of object detectors (Padilla et al., 2020). The measures, precision and recall, are required in the estimation of mAP. Precision is the ratio of true positives to all predicted positives, whereas recall is the ratio of true positives to all actual positives. To explicitly express precision and recall, the following parameters are defined:

- True Positive (TP): If an object or defect instance is present in the ground truth, and the label and the bounding box of the instance are correctly predicted with $\text{IoU} \geq \text{threshold}$.
- False Positive (FP): If the model predicts an object or defect instance at a particular location in the image, but the instance is not present in the ground truth for that particular image. The FP also applies to a

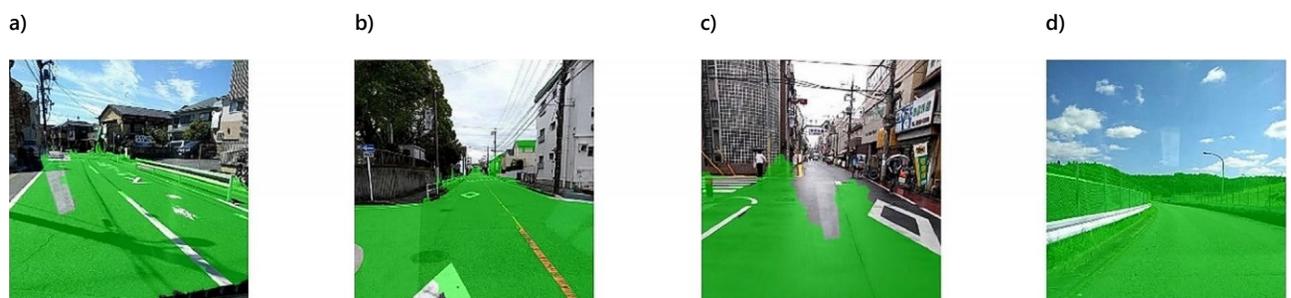


Figure 7. Segmentation trials using the lazy snapping technique on RDD-2020

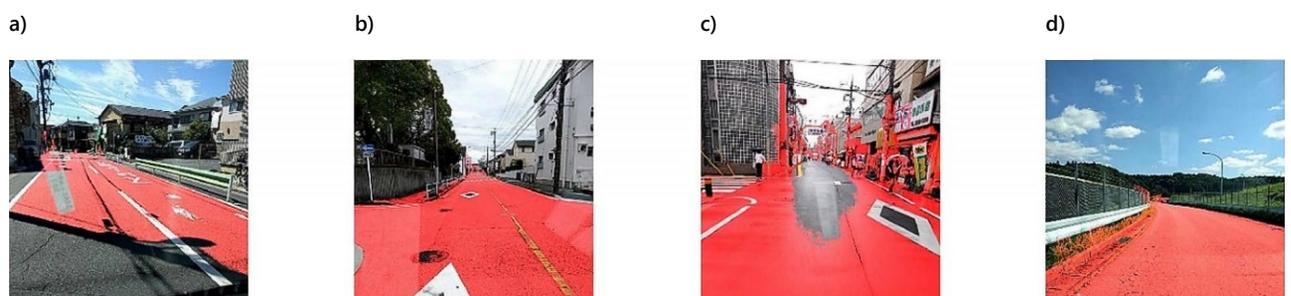


Figure 8. Segmentation trials using region growing from a seed point on the RDD-2020

case in which the predicted label does not match the actual label.

- False Negative (FN): If an object or defect instance is present in the ground truth, but the model fails to predict either the correct label or the bounding box of the instance.

The precision and recall are defined mathematically as:

$$\text{Precision} = \frac{TP}{TP + FP}; \quad (10)$$

$$\text{Recall} = \frac{TP}{TP + FN}. \quad (11)$$

The average precision (AP) is obtained as the average of the precision values obtained from the precision–recall (PR) curve for a select set of recall values. The mAP score is the mean of the APs over all the object classes, N :

$$mAP = \frac{1}{N} \sum_{n=1}^N AP_n. \quad (12)$$

4.4.3. Image annotation

This research explored three annotation cases (Figure 9) with decreasing labor requirements and precision, and compared their accuracy in determining the correct defect classifications and quantifications against expert judgments. The experimentation employed the Visual Geometry Group (VGG) Image Annotator (VIA) software to annotate the RDD-2020 images, and the annotations were exported in the JavaScript Object Notation (JSON) format. The annotation of the objects of interest was done following Table 3. The road feature was added to the defect classes defined by Arya et al. (2020a) with D00, D10, and D20 defining cracks and D40 mainly potholes, based on the Japan Road Maintenance and Repair Guidebook 2013 (JARA, 2013). Figure 10 shows an example image of road damage before (a) and after (b) annotation. Depending on the number of defects that the annotator observed in the image, it took about 2–4 min for case I, 1–2 min for case II, and less than 1 min for case III to annotate a single image using human labor. Figure 11 and Figure 12 show the training and validation data statistics, respectively, with a total of 1,165 annotated objects. The low occurrence of D40 defects in Japan compared to the other countries is probably due to better and more regular maintenance.

4.4.4. Deep learning experiments

The model was trained for 50 epochs at a learning rate of 0.001 in Python 3.9.1 using an Intel® Core™ i5-5200U CPU @ 2.20 GHz with a 4 GB RAM and 500 GB HDD computer. Model accuracy was tested at different IoU thresholds across different object classes considering varied annotation precision. The model training took about 33 h. Table 4 shows the mAP values per defect class at different IoU thresholds and annotation cases. Figure 13 shows the improved results of the road feature extraction and the defect identification done in parallel without image lighting and color change inhibitions (comparing Figures 13a



Figure 9. Different annotation cases in red, green and blue

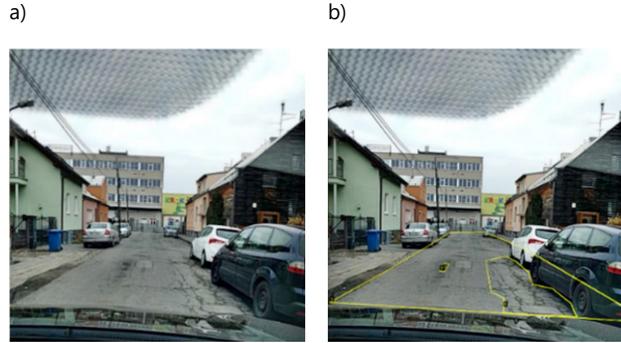


Figure 10. Road damage image before a) and after b) annotation

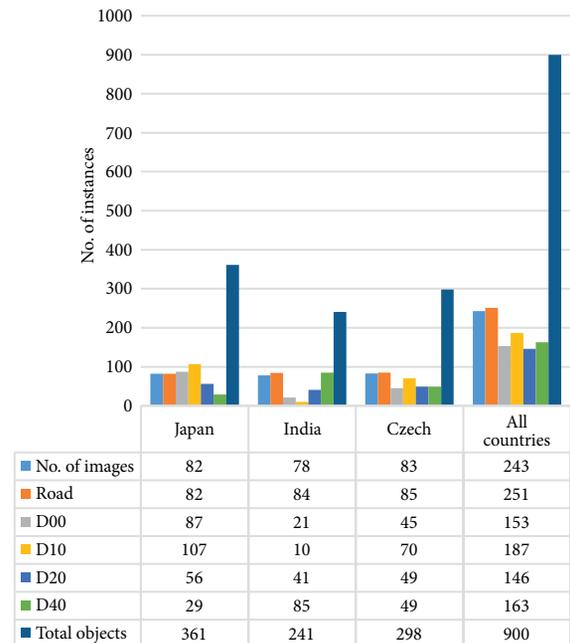


Figure 11. Training data statistics

Table 3. Objects of interest (Arya et al., 2020a)

Object ID	Description	Defects Group
Road	Road surface	
D00	Linear crack, longitudinal	C_k
D10	Linear crack, lateral	
D20	Alligator crack	
D40	Pothole, rutting, bump, separation	D_k

with 8a; 13b with 7b and 8b; and 13c with 7c and 8c) and unwanted segmentation (comparing Figure 13d with 7d). The model showed high confidence values of up to 0.99 for prominent road features.

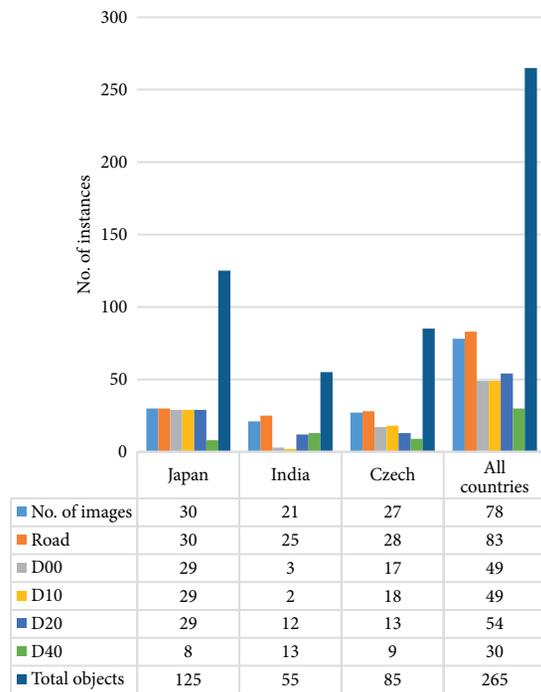


Figure 12. Validation data statistics

Comparing across the different annotation cases, the mAP increased by an average of 8.3% from case I to II and by an average of 5.8% from case II to III considering all IoU thresholds. This mAP increase was due to the increase in the RoI size, which resulted in more overlap between the ground truth and the prediction; hence, more detection. However, the difference in detection accuracy was less than 10% for both comparisons, which may arguably be insignificant. In particular, cases III and II may be competitive because relatively similar mAP levels were achieved at a lower annotation cost compared to case I. This result may show the insignificance of very precise annotations in generating acceptable defect density estimates for pavement management purposes. It is also important to note that for a given annotation case and IoU, specific defects may be better detected; for instance, the smaller size D00 defects were detected at the highest AP considering annotation case II, which is more precise compared to III, for all IoU thresholds.

The Road object class had the highest AP values because the road feature was very prominent in all images, which made it easy for the algorithm to learn, detect, and segment. On the other hand, linear cracks consisting of lateral and longitudinal cracks had comparatively lower APs because they were generally of much smaller size and less prominent compared to other objects; hence, their detection and segmentation was poorer. As the IoU was decreased, the APs increased across all object classes except for the Road class because the less strict IoU requirement resulted in more object detection as the ground truth and the prediction did not need to overlap much. The high AP value for the Road class was stagnant because it was the maximum achievable value.

Table 4. AP per object class at different IoU thresholds

IoU	Object ID	Case I		Case II		Case III	
		AP	mAP	AP	mAP	AP	mAP
0.7	Road	0.9794	0.3436	1.0	0.3963	1.0	0.4097
	D00	0.0833		0.3280		0.1517	
	D10	0.3556		0.0407		0.4520	
	D20	0.1011		0.2045		0.0824	
	D40	0.1985		0.4081		0.3624	
0.5	Road	0.9794	0.5738	1.0	0.6328	1.0	0.6979
	D00	0.3444		0.6472		0.5992	
	D10	0.5799		0.1512		0.7674	
	D20	0.5337		0.5704		0.4843	
	D40	0.4318		0.7951		0.6386	
0.3	Road	0.9794	0.7020	1.0	0.7284	1.0	0.7645
	D00	0.6083		0.7358		0.7258	
	D10	0.5950		0.2523		0.7681	
	D20	0.7291		0.7893		0.6039	
	D40	0.5984		0.8644		0.7245	
0.1	Road	0.9794	0.7046	1.0	0.7309	1.0	0.7645
	D00	0.6083		0.7358		0.7258	
	D10	0.5950		0.2523		0.7681	
	D20	0.7397		0.8019		0.6039	
	D40	0.6005		0.8644		0.7245	

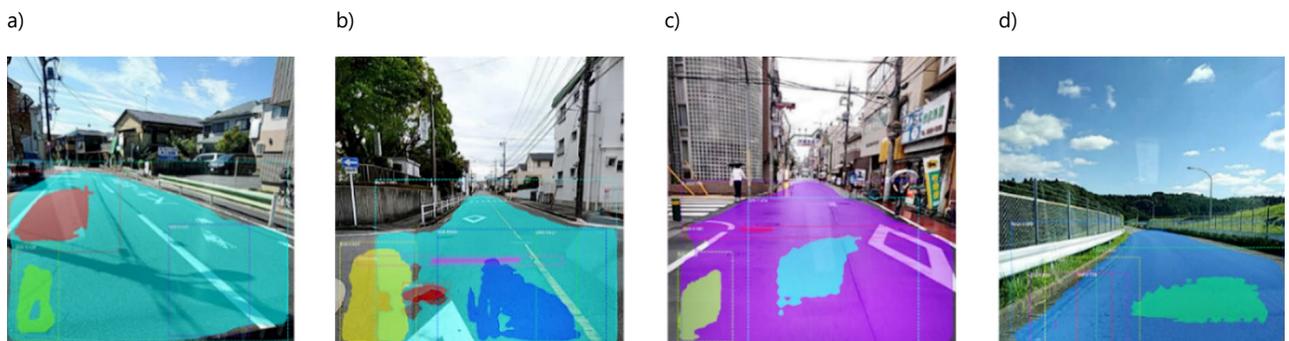


Figure 13. Detection and segmentation of road features and defects on RDD-2020 for case I, IoU 0.5

4.4.5. Defect density

The extent of defects could be estimated from the segmented images by calculating the ratio of the size of defect pixels to the size of pavement pixels. A ratio was considered because the images were taken at different perspectives and using different smartphones and hence detected objects had different relative sizes. This estimation of the defect densities d_k^n in an image taken at a specific location was done so as to facilitate comparisons between different road sections. The defect density is also similar to the cracking ratio defined by the Japan Road Association (JARA, 2013; Kubo, 2017):

$$d_k^n = \frac{\text{No. of defect pixels}}{\text{No. of total pavement pixels}} \quad (13)$$

The defect densities were estimated for 1,660 sections in Adachi City, Japan (Figure 14). For some sections, the aggregate defect densities were greater than expected, probably due to the detection and segmentation of multiple objects at different instances, overlap, and partial detection of road features in the images. From Figure 14, more defects were detected when the IoU was reduced because a less strict IoU enabled more detection. Also, more severe defects could be detected from annotation case I to II, to III; attributable to the increase in the RoI size as annotation was made less precise. The reduction in defect detection for case III could be attributed to annotation variability.

4.5. Estimation of the safety metric

The data preparation stage before application of the probabilistic model generated 12,000 data points for 100 selected sections from Adachi City considering 12 possible combinations of annotation case i and IoU compared against the judgment of ten expert engineers. The experts visually classified the pavement images based on a three-point scale (Table 5) and their experience. To eliminate bias, the experts were presented with the selected images labeled from 1 to 100 (blind judgment). A comparison was made between the expert image classification and the image analysis results. In the estimation of MCI, instead of the rutting/pothole depth, the defect density value was used. The estimated MCI values were compared with the expert analysis to determine whether there was a match or no match.

Table 5. Evaluation of pavement soundness (Kubo, 2017)

Type		Condition (Deterioration level)	Approximate MCI_k
1	Good	Low and the pavement surface is in a good condition	$4.5 \leq MCI_k \leq 10$
2	Phase to keep surface function	Medium deterioration level	$2.5 \leq MCI_k < 4.5$
3	Repair phase	High and expected to be beyond the permissible level soon	$MCI_k < 2.5$

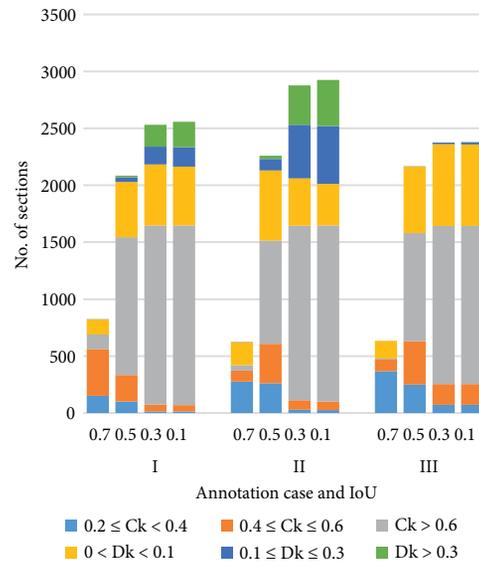


Figure 14. Section defect densities

Table 6 shows a summary of the expert classification aggregated into the worst, best, and mode (majority) results from the ten experts for each of the 100 sections. About 71% of the sections were classified as good by the majority of experts.

Table 6. Aggregate expert classification for the selected sections

Condition state	Number of sections		
	Worst	Best	Mode
1	28	86	71
2	51	13	28
3	21	1	1

Figure 15 shows a comparison of image processing output and expert classification with the highest matching rate of 65.5% achieved for both annotation cases II and III at 0.7 IoU considering all experts.

The unknown parameters were estimated by maximizing the log-likelihood function using Newton's method, as shown in Table 7. All the estimated parameters had significant p -values (< 0.05) and all had low standard errors. The significant p -values mean that i and IoU influence the accurate detection of defects. The estimated parameters β_1 and β_2 were positive showing that matching detection increased as annotation was made less strict for case II and III, respectively, in comparison to the more strict case I. The positive parameters (β_1 and β_2) show that for pavement defect detection tasks, very strict annotations may be unnecessary in determining acceptable pavement condition classifications for effective management and planning decisions. Also, parameter β_3 was positive showing that more matching detections were achieved as the IoU threshold was increased. However, very high IoU thresholds could have led to less defect detection and the classification of more pavements in good condition matching the experts. This result suggests that the IoU threshold should be set

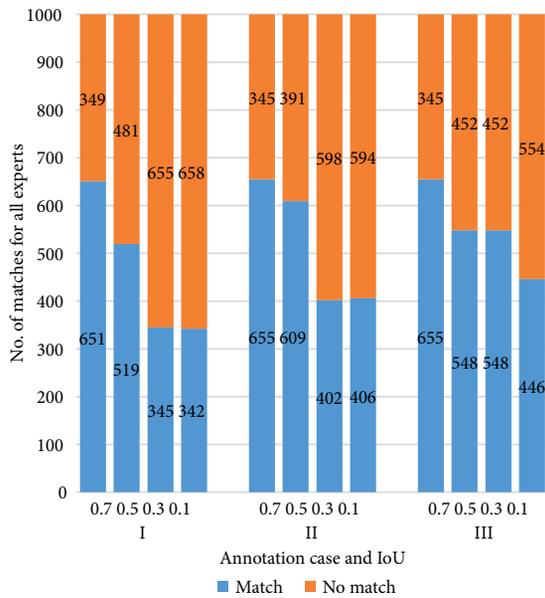


Figure 15. Image processing vs expert classification

high enough to achieve more correct detections but not too high as many defects may be undetected. As the annotation case was made less strict, i.e., from I to II to III, the estimated parameters increased, which showed that the odds of obtaining correct detections increased by 25.2% from I to II, 28.3% from I to III, and 3.1% from II to III, if other variables were fixed. The increase in odds as annotation was made less strict could be attributed to the increase in RoI area, which increased the probability of detection. There were 574% greater odds of correct detection if the IoU requirement was increased and annotation precision was fixed. As discussed above, this increase in odds may also imply that very high IoUs resulted in less defect detection, which generated results that showed road pavements in good condition and matching the expert judgement. These results support the need to limit the IoU within given thresholds to avoid erroneous and meaningless detections as IoU approaches the limits of 0 and 1, respectively.

To evaluate the success rate of the model in making correct predictions, a confusion matrix and hitting rate accuracy were generated. The confusion matrix showing the accuracy rate of the calibrated model in predicting an observed match or no match at a 0.5 cutoff is shown in

Table 8. The model correctly predicted the observations at a 60.42% accuracy rate, showing the high goodness of fit of the calibrated model. The model may not

need to be an exact match since the expert classification was not based on only observed defects. The expert classification was also based on experience and other more informative defect characteristics including crack patterns and defect colors not considered in the deep learning. For example, diagonal cracks are indicative of shear failure in reinforced concrete structures, reflection linear cracks may show the degeneration of flexible pavement sublayers, and a black color may be due to bleeding defects where asphalt binder is forced to the pavement surface.

Table 8. Confusion matrix

		Observed		Total
		Match	No match	
Predicted	Match	3637	2363	6000
	No match	2387	3613	6000
Total		6024	5976	12000
Accuracy		0.6038	0.6046	0.6042

Figure 16 shows a match in classification between the image processing (case II) and the majority of experts with Figure 16a and 16b classified as 1 and 2, respectively, by both the experts and image processing. Figure 17 shows a mismatch in classification between the majority of experts (classified as 3) and case II of image processing (classified as 2) probably attributable to the variation in the grouping of the condition states and other factors such as crack patterns and defect colors not considered in the image processing as discussed above.

Table 9 shows the probabilities of obtaining a match or the correct detection for each *i* and IoU combination. The highest probability of 0.66 was achieved if IoU was set to 0.7 and the least precise annotation case III was applied. As noted, the high probabilities for 0.7 IoU were also influenced by no defect detection, which led to high MCI values that matched the expert results. If p_0 is set to 0.5, then IoUs of 0.3 and below may be inappropriate for pavement defect detection tasks, and high IoUs of 0.7 and above must be selected taking into consideration the inability to detect defects at such strict thresholds. This result showed the insignificance of highly precise annotations (case I) and that very low or very high IoUs may be undesirable for road defect detection tasks. From the optimization process, defect densities can be obtained and proposed section intervention can be determined following an intervention matrix after prioritization.

Table 7. Estimation of unknown parameters

Parameter	Estimate	Standard error	exp(β)	Increase in odds	<i>p</i> -value
β_0	-0.9127	0.04734	0.4014	-0.5986	2×10^{-16}
β_1	0.2250	0.04580	1.2523	0.2523	8.98×10^{-7}
β_2	0.2491	0.04581	1.2829	0.2829	5.42×10^{-8}
β_3	1.9074	0.08483	6.7356	5.7356	2×10^{-16}



Figure 16. Image processing and expert classification match



Figure 17. Image processing and expert classification no match

Table 9. Probabilities of detecting a match

Annotation case i	IoU	Pr (i , IoU)
I	0.7	0.6041
	0.5	0.5102
	0.3	0.4157
	0.1	0.3270
II	0.7	0.6564
	0.5	0.5661
	0.3	0.4712
	0.1	0.3783
III	0.7	0.6619
	0.5	0.5720
	0.3	0.4772
	0.1	0.3839

3.6. Intervention prescription

Intervention was prescribed based on the estimated safety metric for the 100 sections classified by both the experts and through image processing. The MCI_k mean values (shown in Appendix, Table 11) show decreasing MCI (worsening pavement condition) as IoU was made less strict for all annotation cases because more defects were detected at lower IoU thresholds. Also, the standard deviations had less difference as annotation was made less strict across IoUs of 0.5 and below due to the reduced variability in annotation quality.

The sections selected for intervention based on MCI_k^* optimization and the intervention matrix in Table 1, with MCI_k set to 0 and considering a limitless budget, for each annotation case at 0.5 IoU are shown in Table 10. The results showed similarity between annotation case II and III as all the 100 sections were selected for either crack sealing or patching and overlay with a bigger proportion proposed for crack sealing or patching. The difference for annotation case I could be attributable to higher annotation variability.

Table 10. Number of sections selected for intervention

Annotation case i	Intervention			
	A_0	A_1	A_2	A_3
I	5	33	61	1
II	0	80	20	0
III	0	59	41	0

5. Discussion

5.1. Annotation precision and cost trade-off

This study evaluated the trade-off between image processing inputs including annotation precision and cost. First, the research experiment explored the use of simple segmentation methods that do not require costly annotation on the publicly available RDD-2020. For efficiency reasons, simple segmentation should be done programmatically as opposed to manually since the latter is cumbersome for a large image dataset. Although simple segmentation methods do not incur annotation costs, they were challenged by region breakage due to poor lighting conditions and color changes, and unwanted objects were segmented in several cases. These challenges could have potentially resulted in inaccurate estimation of defect densities.

The shortcomings of simple segmentation methods led to experimenting on deep learning methods. A deep learning model was trained using the Mask R-CNN algorithm utilizing the RDD-2020. For deep learning, the road images were annotated by varying the degree of precision. Less informative annotations required less annotation time, whereas more informative annotations took more time. The difference in the mAP values for the different annotation cases were arguably insignificant. Additionally, as the annotation precision was reduced, the odds of obtaining a correct match increased from annotation case I, II to III as a result of the increase in the RoI area, which increased the probability of defect detection. These results showed the insignificance of very precise and costly annotations for pavement defect detection tasks. However,

the annotations should be tight enough to avoid erroneous allocation of non-object regions as RoIs before model training as this may generate inaccuracies.

5.2. Deep learning, objective annotation, and IoU

As discussed in the previous subsection, deep learning overcame the challenges faced by simple segmentation methods. In addition, deep learning in which all objects of interest (road features and defects) were identified and segmented in parallel made it convenient to estimate the defect densities. The deep learning results were also more promising with objects detected and segmented with high AP values. Road features were detected and segmented with the highest AP values because they were very prominent in the images. As the IoU threshold was decreased, more defects were detected because a less strict IoU requirement enabled more detection since the ground truth and the prediction did not need to overlap much. Also, more severe defects were detected from annotation case I to II to III because of the increase in the RoI size, which allowed more overlap between the ground truth and the prediction.

A probabilistic pavement management model that included setting the IoU and annotation precision objectively with validation from experts was developed in this study to facilitate the standardization of setting model inputs to have more uniform outputs that would minimize the variability of infrastructure intervention decisions. Validation of the model by a group of experts ensured that the model could be applied even in areas where there were very few or no expert engineers, and hopefully reduce the proportion of infrastructure neglected as a result of personnel shortages. In the empirical application of the model, the estimated parameters were significant, which showed that the IoU threshold and annotation precision influenced accurate defect detection. The calibrated model also had a high success hitting rate of more than 60%. It was shown that a low IoU resulted in more defect detection, whereas the annotation precision was insignificant. As the annotations were made less precise, the RoI size increased, which resulted in more overlap between the ground truth and the prediction, and hence more defect detection. As the IoU was reduced, the requirement for the area of overlap was made less strict, which meant that the ground truth and the prediction did not need to overlap much, increasing detection. However, the IoU threshold must be restricted to avoid meaningless and erroneous detections if the IoU approaches the limits of 0 and 1. Despite being insignificant, the object annotations must be as tight as possible to avoid labeling non-object regions as objects, and the annotations should not be cumbersome so as to lower the processing cost. The developed model maximized the probability of correct defect detection because more detection may encourage proactive intervention and further investigation for the candidate sections, which could improve road user safety.

5.3. Road asset management application

The optimum model inputs, annotation precision and IoU, could be used to generate section defect densities and safety level, MCI, in this paper. The obtained MCI that may closely match the classification by experts can then be used to prescribe the appropriate interventions for a road pavement group using MCI optimization and an intervention matrix, as shown earlier. The intervention on a section can be based on the severity of the defects on a given section in comparison to other sections in the infrastructure group. It may be noted that the intervention in this case could be based on one-point image data collected at one time point; however, the appropriateness of intervention decisions could be improved by incorporating future performance prediction models such as the stochastic Markov hazard model since more image data is expected to become available in the future. The required performance model could inform the image data collection process so as to generate more consistent and usable data for pavement performance prediction.

6. Conclusions

This research proposed a framework to feasibly apply deep learning model results to pavement asset management. The study used publicly available smartphone road image data from Japan, India, and the Czech Republic to train and validate a deep learning model built on the Mask R-CNN algorithm. The experiments showed that with fewer management resources for an increasing infrastructure stock, computer vision promises safer and more efficient asset management and planning compared to the current human-dependent practice. The research work empirically showed the following:

- 1) Experimental comparison showed the merits of deep learning compared to simple segmentation in overcoming poor lighting conditions and color changes to correctly segment objects in images.
- 2) Choice of the IoU threshold and annotation precision are important for object detection tasks and should be optimally determined. The IoU significantly affected defect detection and hence should be carefully selected to avoid meaningless and erroneous detections. As the annotation precision was insignificant in defect detection, less costly simplified polygons may suffice.
- 3) Tests showed that one-point data obtained from a single image dataset can be efficiently used to support intervention choices on infrastructure with less human dependence.

In the future, better stochastic asset management models could be developed using consistently obtained data because the current state-of-the-art asset management models require at least two-point condition data to model deterioration processes and perform LCC analysis. Building and improving algorithms that detect the patterns of defects and their colors is a possible area for future re-

search because defect patterns and colors could indicate the failure type. The efficiency of detection and segmentation algorithms also needs to be generally improved. A positive feedback loop could also be created between data collection and future asset management needs. To make data more usable for planning, the global positioning system (GPS) coordinates of photos could be included in databases so that road sections can be better identified and linked to road network maps. It is recommended that further studies and methodologies be developed to make annotations more objective than subjective and thus minimize the variability due to different annotators.

Acknowledgements

The authors appreciate Sekimoto Laboratory at the University of Tokyo for making the road damage dataset publicly available and the expert engineers from PASCO Corporation for grading the road pavement images.

Funding

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI [Grant Number: 21K04288].

Author contributions

Felix Obunguta conceived the study, wrote the article, wrote and modified code used in the study, carried out data analysis and review. Kakuya Matsushima sharpened the conceptual ideas in the article, supervised the production and reviewed the article. Junichi Susaki contributed ideas, reviewed the article and code used.

Disclosure statement

The authors have no competing financial, professional, or personal interests from any parties.

References

- Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., Mraz, A., Kashi-yama, T., & Sekimoto, Y. (2020a). Transfer learning-based road damage detection for multiple countries. arXiv:2008.13101. <https://doi.org/10.48550/arXiv.2008.13101>
- Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., Mraz, A., Kashi-yama, T., & Sekimoto, Y. (2020b). Global road damage detection: state-of-the-art solutions. In *IEEE International Conference on Big Data (Big Data)* (pp. 5533–5539). <https://doi.org/10.1109/BigData50022.2020.9377790>
- Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., Mraz, A., Kashi-yama, T., & Sekimoto, Y. (2021). Deep learning-based road damage detection and classification for multiple countries. *Automation in Construction*, 132, Article 103935. <https://doi.org/10.1016/j.autcon.2021.103935>
- Barchard, K. A., & Pace, L. A. (2011). Preventing human error: The impact of data entry methods on data accuracy and statistical results. *Computers in Human Behavior*, 27(5), 1834–1839. <https://doi.org/10.1016/j.chb.2011.04.004>
- Bosurgi, G., Modica, M., Pellegrino, O., & Sollazzo, G. (2022). An automatic pothole detection algorithm using pavement 3D data. *International Journal of Pavement Engineering*. <https://doi.org/10.1080/10298436.2022.2057978>
- Cambridge University. (2021). *CamVid dataset*. <http://mi.eng.cam.ac.uk/research/projects/VideoRec/CamVid/>
- Deng, J., Dong, W., Socher, R., Li, L. -J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 248–255), Florida, USA. <https://doi.org/10.1109/CVPR.2009.5206848>
- Everingham, M., Eslami, S. A., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2015). The Pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1), 98–136. <https://doi.org/10.1007/s11263-014-0733-5>
- Georgopoulos, A., Loizos, A., & Flouda, A. (1995). Digital image processing as a tool for pavement distress evaluation. *ISPRS Journal of Photogrammetry and Remote sensing*, 50(1), 23–33. [https://doi.org/10.1016/0924-2716\(95\)91844-A](https://doi.org/10.1016/0924-2716(95)91844-A)
- Girshick, R. (2015). Fast R-CNN. In *IEEE International Conference on Computer Vision* (pp. 1440–1448), Boston, MA, USA. <https://doi.org/10.1109/ICCV.2015.169>
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 580–587), Ohio, USA. <https://doi.org/10.1109/CVPR.2014.81>
- Goncalves, L. R., & Givigi, S. N. (2016). Automatic crack detection and measurement based on image analysis. *IEEE Transactions on Instrumentation and Measurement*, 65(3), 583–590. <https://doi.org/10.1109/TIM.2015.2509278>
- Greenwald, N. F., Miller, G., Moen, E., Kong, A., Kagel, A., Dougherty, T., Fullaway, C. C., McIntosh, B. J., Leow, K. X., Schwartz, M. S., Pavelchek, C., Cui, S., Camplisson, I., Bar-Tal, O., Singh, J., Fong, M., Chaudhry, G., Abraham, Z., Moseley, J., ... Van Valen, D. (2022). Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nature Biotechnology*, 40, 555–565. <https://doi.org/10.1038/s41587-021-01094-0>
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2018). *Mask R-CNN, Facebook AI Research (FAIR)*. arXiv:1703.06870. <https://doi.org/10.48550/arXiv.1703.06870>
- Hong, J. W., Jin, S., & Lee, S. E. (2020). A vision-based approach for autonomous crack width measurement with flexible kernel. *Automation in Construction*, 110, Article 103019. <https://doi.org/10.1016/j.autcon.2019.103019>
- JARA. (2013). *Maintenance and repair guide book of the pavement 2013* (1st ed.). Japan Road Association, Tokyo.
- Kerali, H. G. (2000). *HDM-4: Highway development and management. Volume one: overview of HDM-4*.
- Kinyanjui, N. M., Odonga, T., Cintas, C., Codella, N. C. F., Panda, R., Sattigeri, P., & Varshney, K. R. (2019). Estimating skin tone and effects on classification performance in dermatology datasets. In *NeurIPS Workshop on Fair ML for Health*, Vancouver, Canada. <https://doi.org/10.48550/arXiv.1910.13268>
- Kobayashi, K., Do, M., & Han, D. (2010). Estimation of Markovian transition probabilities for pavement deterioration forecasting. *KSCSE Journal of Civil Engineering*, 14(3), 343–351. <https://doi.org/10.1007/s12205-010-0343-x>
- Kobayashi, K., Kaito, K., & Lethan, N. (2012). A Bayesian estimation method to improve deterioration prediction for infrastructure system with Markov chain model. *International Journal of Architecture, Engineering and Construction*, 1(1), 1–13. <https://doi.org/10.7492/IJAEC.2012.001>

- Kobayashi, K., Eguchi, M., Oi, A., Aoki, K., & Kaito, K. (2013). The optimal implementation policy for inspecting pavement with deterioration uncertainty. *Journal of Japan Society of Civil Engineers*, 1(1), 551–568. https://doi.org/10.2208/journalofjsce.1.1_551
- Kroon, D. (2021). *Region growing*. MATLAB Central File Exchange. <https://www.mathworks.com/matlabcentral/fileexchange/19084-region-growing>
- Kubo, K. (2017). Pavement maintenance in Japan. In *Road Conference International Symposium*. https://www.road.or.jp/international/pdf/32_AM6.pdf
- Lethanh, N., & Adey, B. T. (2012). A hidden Markov model for modeling pavement deterioration under incomplete monitoring data. *International Journal of Civil and Environmental Engineering*, 6(1). <https://doi.org/10.5281/zenodo.1082516>
- Lin, K., & Lin, C. (2011). Applying utility theory to cost allocation of pavement maintenance and repair. *International Journal of Pavement Research and Technology*, 4(4), 212–221.
- Liu, H., & Wang, D. Z. W. (2016). Modeling and solving discrete network design problem with stochastic user equilibrium. *Journal of Advanced Transportation*, 50(7), 1295–1313. <https://doi.org/10.1002/atr.1402>
- Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T., & Omata, H. (2018). Road damage detection and classification using deep neural networks with smartphone images. *Journal of Computer-Aided Civil and Infrastructure Engineering*, 33, 1127–1141. <https://doi.org/10.1111/mice.12387>
- Minami, M., & Suzuki, T. (2008). Pavement maintenance level and annual budget over the regional road network. *Journal of Construction Management*, 15, 71–79. <https://doi.org/10.2208/procm.15.71>
- Mirikharaji, Z., Abhishhek, K., Izadi, S., & Hamarneh, G. (2021). D-LEMA: Deep learning ensembles from multiple annotations application to skin lesion segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops* (pp. 1837–1846). <https://doi.org/10.1109/CVPRW53098.2021.00203>
- Miyamoto, A., & Yoshitake, T. (2009). Development of a remote condition assessment system for road infrastructures. In *International ECCE Conference, EUROINFRA 2009, Current State and Challenges for Sustainable Development of Infrastructure*. <https://www.irbnet.de/daten/iconda/CIB17687.pdf>
- Mizutani, D., Nakazato, Y., & Lee, J. (2020). Network-level synchronized pavement repair and work zone policies: Optimal solution and rule-based approximation. *Transportation Research Part C: Emerging Technologies*, 120, Article 102797. <https://doi.org/10.1016/j.trc.2020.102797>
- Obunguta, F., & Matsushima, K. (2020). Optimal pavement management strategy development with a stochastic model and its practical application to Ugandan national roads. *International Journal of Pavement Engineering*, 23(7), 2405–2419. <https://doi.org/10.1080/10298436.2020.1857759>
- Obunguta, F., Matsushima, K., & Bakamwesiga, H. (2022). Social cost optimization model and empirical evaluation of intervention effects on Ugandan road pavements. *Journal of Infrastructure Systems*, 28(4), Article 05022005. [https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000707](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000707)
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62–66. <https://doi.org/10.1109/TSMC.1979.4310076>
- Padilla, R., Netto, S. L., & da Silva, E. A. B. (2020). A survey on performance metrics for object-detection algorithms. In *Proceedings of the International Conference on Systems, Signals and Image Processing (IWSSIP)* (pp. 237–242). <https://doi.org/10.1109/IWSSIP48289.2020.9145130>
- Pérez-Acebo, H., Mindrab, N., Railean, A., & Rojí, E. (2019). Rigid pavement performance models by means of Markov Chains with half-year step time. *International Journal of Pavement Engineering*, 20, 830–843. <https://doi.org/10.1080/10298436.2017.1353390>
- Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. arXiv:1506.01497. <https://doi.org/10.48550/arXiv.1506.01497>
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63, 581–592. <https://doi.org/10.1093/biomet/63.3.581>
- Rubin, D. B. (1987). *Multiple imputation for non-response in surveys*. John Wiley. <https://doi.org/10.1002/9780470316696>
- Tabatabaee, N., & Ziyadi, M. (2013). Bayesian approach to updating Markov-based models for predicting pavement performance. *Transportation Research Record: Journal of the Transportation Research Board*, 2366(1), 34–42. <https://doi.org/10.3141/2366-04>
- Thao, N. D., Aoki, K., Kato, T., Toan, T. N., Kobayashi, K., & Kaito, K. (2015). A practical process to introduce a customized pavement management system in Vietnam. *Journal of Japan Society of Civil Engineers*, 3(1), 246–258. https://doi.org/10.2208/journalofjsce.3.1_246
- The MathWorks Inc. (2021). <https://uk.mathworks.com/help/images/ref/lazysnapping.html>
- Thuyet, D. Q., Jomoto, M., Hirakawa, K., & Lei Swe, Y. L. (2022). Development of an autonomous road surface damage inspection program using deep convolutional neural network. *Journal of Japan Society of Civil Engineers*, 10(1), 235–246. https://doi.org/10.2208/journalofjsce.10.1_235
- Tsuda, T., Kaito, K., Aoki, K., & Kobayashi, K. (2006). Estimating Markovian transition probabilities for bridge deterioration forecasting. *Structural Engineering/Earthquake Engineering*, 23(2), 241s–256s. <https://doi.org/10.2208/jsceseee.23.241s>
- Xu, Y., Zhong, Z., Lian, D., Li, J., Li, Z., Xu, X., & Gao, S. (2021). Crowd counting with partial annotations in an image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 15570–15579). <https://doi.org/10.1109/ICCV48922.2021.01528>
- Yoshida, T. (2016). Composite indicators for assessing maintenance needs for road pavements from the view point of road functions. *Journal of Japan Society of Civil Engineers, Ser. E1 (Pavement Engineering)*, 72(1), 12–20. <https://doi.org/10.2208/jscejpe.72.12>
- Zou, D., Zhang, M., Bai, Z., Liu, T., Zhou, A., Wang, X., Cui, W., & Zhang, S. (2022). Multicategory damage detection and safety assessment of post-earthquake reinforced concrete structures using deep learning. *Journal of Computer-Aided Civil and Infrastructure Engineering*, 37(9), 1188–1204. <https://doi.org/10.1111/mice.12815>

APPENDIX

The Table 11 shows MCI_k values for each annotation case i and IoU for the 100 sections from Adachi City obtained through image processing and majority expert classification.

Table 11. MCI_k values for each annotation case and IoU for the 100 sections and majority expert classification

Section k	Annotation case i and IoU (Image processing)												Majority expert grade
	I				II				III				
	0.7	0.5	0.3	0.1	0.7	0.5	0.3	0.1	0.7	0.5	0.3	0.1	
Adachi_20170906094007.jpg	10.0	4.29	2.72	2.72	10.0	5.07	3.16	2.54	10.0	4.25	3.46	3.46	1
Adachi_20170906094019.jpg	10.0	4.65	3.33	3.33	5.59	5.59	5.59	5.59	10.0	4.51	3.71	3.71	1
Adachi_20170906102859.jpg	5.92	3.70	3.10	3.10	10.0	5.73	3.13	3.13	5.97	4.55	2.44	2.44	2
Adachi_20170906102924.jpg	8.43	3.56	3.45	2.99	6.93	6.15	2.51	2.51	10.0	3.75	3.45	3.45	2
Adachi_20170906103322.jpg	5.62	4.38	2.93	2.93	10.0	5.29	3.52	3.52	8.14	4.00	3.41	3.41	1
Adachi_20170906105116.jpg	10.0	3.71	2.68	2.68	10.0	8.46	3.63	3.63	6.22	4.75	4.75	4.75	1
Adachi_20170906105131.jpg	10.0	5.56	4.61	4.61	7.41	6.47	5.43	5.43	10.0	4.29	4.11	4.11	1
Adachi_20170906105143.jpg	10.0	4.31	2.55	2.55	10.0	5.88	3.68	3.68	6.40	3.64	3.39	3.39	1
Adachi_20170906143942.jpg	10.0	10.0	3.27	3.27	10.0	4.45	2.48	2.48	6.68	4.21	3.42	3.42	1
Adachi_20170906143948.jpg	10.0	9.93	4.08	4.08	10.0	9.90	4.24	4.24	10.0	4.20	4.20	4.20	1
Adachi_20170906144435.jpg	8.27	3.76	3.44	3.44	6.25	2.80	2.80	2.80	6.11	5.05	3.82	3.82	1
Adachi_20170906150612.jpg	10.0	3.48	3.31	3.31	10.0	4.31	2.52	2.52	5.71	4.32	3.84	3.84	1
Adachi_20170906151235.jpg	10.0	3.07	2.46	2.46	10.0	2.85	2.74	2.64	6.43	4.06	3.65	3.65	2
Adachi_20170906151255.jpg	8.51	3.81	3.81	3.81	10.0	5.55	3.30	3.30	10.0	4.39	3.89	3.89	2
Adachi_20170906153240.jpg	10.0	4.79	3.90	3.90	10.0	4.30	4.30	4.30	5.33	4.62	4.02	4.02	1
Adachi_20170906153353.jpg	10.0	5.37	4.89	4.89	5.00	4.45	4.63	4.63	6.77	6.08	5.23	5.23	1
Adachi_20170906153805.jpg	5.83	3.98	3.41	3.41	5.54	3.87	3.87	3.87	6.01	5.40	3.94	3.94	1
Adachi_20170906154228.jpg	10.0	10.0	3.28	3.28	7.17	6.18	3.27	3.27	7.87	5.19	4.00	4.00	1
Adachi_20170906154236.jpg	10.0	5.84	5.84	5.84	10.0	5.40	2.82	2.82	5.75	5.07	4.49	4.49	1
Adachi_20170906155248.jpg	6.91	3.42	2.55	2.55	10.00	5.30	3.51	3.51	9.80	4.34	3.34	3.34	1
Adachi_20170906155457.jpg	10.0	3.81	3.81	3.81	10.0	5.97	3.77	3.77	10.0	5.76	4.69	4.69	1
Adachi_20170906160907.jpg	10.0	6.18	5.46	5.46	5.79	4.84	3.44	3.44	10.0	5.62	5.38	5.38	1
Adachi_20170907134021.jpg	5.12	3.42	3.24	3.24	6.10	3.71	3.71	3.71	5.92	3.85	3.85	3.85	2
Adachi_20170907134352.jpg	10.0	3.58	2.89	2.89	6.04	4.68	3.08	3.08	5.77	4.22	3.59	3.59	1
Adachi_20170907134444.jpg	10.0	3.26	3.26	3.26	7.07	6.53	3.70	3.70	10.0	4.75	4.70	4.70	1
Adachi_20170907134447.jpg	10.0	4.61	3.83	3.83	10.0	5.39	3.04	3.04	7.34	4.48	4.04	4.04	1
Adachi_20170907134508.jpg	10.0	4.45	3.45	3.45	10.0	5.03	3.50	3.50	10.0	5.67	5.67	5.67	1
Adachi_20170907135021.jpg	9.91	4.93	3.12	3.12	8.93	5.42	3.21	3.21	6.32	5.13	5.13	5.13	1
Adachi_20170907140937.jpg	10.0	4.37	3.35	3.35	10.0	6.18	4.11	4.11	5.38	4.78	3.93	3.93	1
Adachi_20170907142138.jpg	10.0	5.85	4.37	4.37	9.89	5.98	4.12	4.12	10.0	5.25	5.25	5.25	1
Adachi_20170907142141.jpg	10.0	5.53	2.50	2.29	6.31	5.86	3.92	3.92	5.03	4.57	4.57	4.57	1
Adachi_20170907142609.jpg	5.12	4.97	3.61	3.61	10.0	6.20	3.65	3.49	6.03	4.82	4.92	4.92	2
Adachi_20170907145734.jpg	5.18	4.35	3.79	3.79	6.90	5.48	4.36	4.36	6.12	4.68	4.68	4.68	2
Adachi_20170907150258.jpg	10.0	10.0	10.0	10.0	9.87	9.87	9.87	9.87	6.05	5.38	5.38	5.38	1
Adachi_20170907151702.jpg	10.0	10.0	2.97	2.97	10.0	9.11	4.90	4.90	5.63	4.94	4.44	4.44	1
Adachi_20170907152141.jpg	7.35	3.86	2.32	2.32	6.50	6.50	4.37	4.37	6.06	3.68	3.43	3.43	1
Adachi_20170907152148.jpg	10.0	10.0	2.52	2.52	6.30	5.88	4.64	4.64	6.21	4.79	3.92	3.86	1
Adachi_20170908093837.jpg	9.87	4.54	2.58	2.58	6.57	6.57	3.10	3.10	10.0	4.82	3.49	3.49	1
Adachi_20170908100005.jpg	8.88	3.01	2.54	2.54	9.89	5.39	2.82	2.82	6.13	4.26	3.69	3.69	1

Continuation of Table 11

Section <i>k</i>	Annotation case <i>i</i> and IoU (Image processing)												Majority expert grade
	I				II				III				
	0.7	0.5	0.3	0.1	0.7	0.5	0.3	0.1	0.7	0.5	0.3	0.1	
Adachi_20170908100119.jpg	9.72	3.85	3.01	3.01	10.0	5.59	2.99	2.94	6.58	4.25	3.48	3.48	2
Adachi_20170908100151.jpg	7.50	3.06	3.06	3.06	7.88	3.87	2.70	2.70	5.79	3.71	3.54	3.54	2
Adachi_20170908100920.jpg	6.15	4.57	2.68	2.68	10.0	6.42	5.15	5.15	5.93	4.20	3.87	3.87	1
Adachi_20170908100923.jpg	4.98	2.99	2.98	2.98	5.90	5.32	3.88	3.88	5.47	3.97	3.85	3.85	2
Adachi_20170908100939.jpg	8.89	3.66	3.66	3.66	6.57	4.71	3.11	3.11	10.0	5.01	5.01	5.01	1
Adachi_20170908100951.jpg	10.0	4.28	4.22	4.22	7.95	4.70	2.98	2.98	8.09	5.07	5.07	5.07	1
Adachi_20170908101018.jpg	8.29	4.12	4.12	4.12	10.0	4.45	4.45	4.45	10.0	4.59	4.42	4.42	2
Adachi_20170908101024.jpg	10.0	3.95	3.95	3.95	6.18	6.18	4.09	3.39	6.42	4.50	4.36	4.36	2
Adachi_20170908101431.jpg	8.23	5.74	4.40	4.40	10.0	6.53	3.95	3.95	10.0	4.83	4.74	4.74	2
Adachi_20170908103050.jpg	10.0	3.96	2.65	2.65	10.0	5.84	3.27	3.01	10.0	4.36	4.36	4.36	2
Adachi_20170908103330.jpg	8.90	4.40	2.68	2.68	8.86	8.86	4.15	4.08	6.55	5.40	4.42	4.42	2
Adachi_20170908103708.jpg	10.0	3.68	3.68	3.68	9.03	6.54	3.95	3.95	6.63	4.96	3.95	3.95	1
Adachi_20170908103711.jpg	10.0	4.24	3.79	3.79	10.0	5.85	4.62	4.62	5.51	4.59	3.57	3.57	1
Adachi_20170908104001.jpg	7.25	3.21	2.37	2.37	10.0	5.27	3.88	3.88	4.82	4.34	3.09	3.09	2
Adachi_20170908104028.jpg	5.43	2.32	1.66	1.45	6.55	2.81	2.25	2.25	5.04	3.27	3.13	3.13	3
Adachi_20170908104341.jpg	8.47	5.15	5.15	5.15	9.75	6.32	3.64	3.64	9.79	5.17	5.17	5.17	1
Adachi_20170908104437.jpg	7.45	4.24	3.61	3.61	10.0	5.59	3.43	3.43	5.61	4.30	3.96	3.96	1
Adachi_20170908104502.jpg	10.0	4.61	4.40	4.40	6.47	4.69	3.36	3.13	7.71	6.62	5.04	5.04	1
Adachi_20170908104520.jpg	7.54	2.99	2.85	2.85	6.97	4.31	3.17	3.17	5.70	3.40	2.82	2.82	1
Adachi_20170908135014.jpg	10.0	5.62	5.62	5.62	5.59	5.59	3.94	3.94	10.0	5.18	4.57	4.57	1
Adachi_20170908141637.jpg	10.0	7.46	6.02	6.02	9.81	5.43	4.49	4.49	10.0	5.94	5.94	5.94	1
Adachi_20170908142026.jpg	10.0	2.62	2.62	2.62	6.43	6.20	5.07	5.07	10.0	5.33	4.42	4.42	1
Adachi_20170908142138.jpg	10.0	4.51	3.57	2.37	10.0	6.20	4.59	4.59	5.81	4.42	3.46	3.46	1
Adachi_20170908144014.jpg	10.0	3.96	1.95	1.95	10.0	5.71	3.33	3.33	6.55	5.49	4.90	4.90	2
Adachi_20170908144258.jpg	7.68	4.40	3.48	3.48	10.0	6.16	4.29	3.48	7.90	6.04	5.54	5.54	2
Adachi_20170908144401.jpg	5.77	3.03	3.03	3.03	10.0	6.17	4.25	4.25	6.53	4.59	3.96	3.96	2
Adachi_20170908144724.jpg	5.85	3.16	3.10	3.10	10.0	6.18	4.59	4.59	6.30	4.07	4.07	4.07	1
Adachi_20170908150714.jpg	9.85	6.39	2.91	2.91	7.32	4.15	4.05	4.05	9.84	4.23	4.23	4.23	1
Adachi_20170911101139.jpg	10.0	5.01	3.88	3.88	6.26	6.20	2.96	2.96	5.14	4.07	3.40	3.40	1
Adachi_20170911101331.jpg	10.0	3.28	2.62	2.62	10.0	3.71	2.79	2.79	5.22	3.68	3.68	3.68	1
Adachi_20170911101544.jpg	4.89	4.13	3.03	3.03	5.08	2.78	2.78	2.78	6.28	3.62	2.90	2.90	1
Adachi_20170911102454.jpg	10.0	3.92	3.49	3.49	7.07	3.91	3.91	3.66	10.0	5.19	4.71	4.71	1
Adachi_20170911103837.jpg	4.49	3.91	3.87	3.87	10.0	8.88	3.81	3.81	5.36	3.79	3.71	3.71	1
Adachi_20170911112332.jpg	4.79	3.54	3.24	3.24	9.67	4.72	4.39	4.39	9.82	5.54	5.54	5.54	1
Adachi_20170911112605.jpg	10.0	4.29	4.29	4.29	10.0	3.88	3.42	3.42	6.40	5.03	5.03	5.03	2
Adachi_20170912093442.jpg	8.77	5.66	4.02	4.02	9.52	4.45	3.62	3.62	7.39	5.43	3.39	3.39	2
Adachi_20170912152741.jpg	5.60	4.07	3.16	3.16	5.25	4.47	3.35	3.04	6.16	4.34	3.85	3.85	2
Adachi_20170913104249.jpg	5.97	5.97	4.09	4.09	9.92	5.60	3.59	3.59	5.79	4.10	4.07	4.07	1
Adachi_20170913104408.jpg	6.13	3.72	3.72	3.72	6.21	6.00	3.68	3.68	6.31	4.70	4.70	4.70	2
Adachi_20170913105158.jpg	5.68	3.31	2.82	2.82	6.39	5.28	3.22	3.22	5.93	4.31	3.59	3.59	2
Adachi_20170913112957.jpg	6.27	3.80	1.85	1.85	6.65	4.68	4.68	4.68	10.0	4.60	4.36	4.36	2
Adachi_20170913113152.jpg	10.0	5.43	3.36	3.36	10.0	7.20	4.17	4.17	10.0	4.75	4.75	4.75	1
Adachi_20170913113542.jpg	6.16	3.87	2.71	2.71	7.18	6.48	3.82	3.82	6.32	4.62	4.62	4.62	1

End of Table 11

Section <i>k</i>	Annotation case <i>i</i> and IoU (Image processing)												Majority expert grade
	I				II				III				
	0.7	0.5	0.3	0.1	0.7	0.5	0.3	0.1	0.7	0.5	0.3	0.1	
Adachi_20170914105125.jpg	10.0	9.91	3.96	3.96	10.0	8.42	5.92	5.92	9.81	5.91	4.57	4.57	1
Adachi_20170914111358.jpg	10.0	4.33	4.33	4.33	10.0	7.25	5.34	5.34	10.0	4.99	4.95	4.95	1
Adachi_20170914141906.jpg	10.0	9.79	4.15	4.15	8.18	6.46	6.46	6.46	10.0	4.87	4.31	3.42	1
Adachi_20170914150611.jpg	5.44	3.49	3.49	3.49	10.0	5.86	2.89	2.89	6.35	4.07	4.07	4.07	1
Adachi_20170914151446.jpg	10.0	4.59	3.68	3.68	10.0	7.43	7.28	7.28	10.0	5.68	3.68	3.68	1
Adachi_20170914155050.jpg	5.11	3.84	3.84	3.84	6.51	6.51	3.99	3.99	10.0	3.60	2.80	2.80	1
Adachi_20170920141001.jpg	10.0	8.89	4.88	4.88	10.0	5.33	3.30	3.30	7.15	3.74	3.47	3.47	1
Adachi_20170920144854.jpg	9.84	5.13	4.63	4.63	9.84	7.19	5.07	5.07	6.71	5.45	5.45	5.45	1
Adachi_20170920144855.jpg	9.85	4.70	4.44	4.44	7.37	5.29	4.20	4.20	6.08	5.31	4.58	4.58	1
Adachi_20170920150338.jpg	4.67	3.76	3.11	3.11	10.0	5.30	4.47	4.47	10.0	4.44	3.48	3.48	2
Adachi_20170920150551.jpg	10.0	3.90	3.68	3.68	5.85	5.85	4.38	4.38	10.0	6.66	3.18	3.18	1
Adachi_20170920150559.jpg	5.06	4.05	3.67	3.67	5.32	5.31	4.38	4.38	10.0	5.84	4.71	4.71	1
Adachi_20170920150923.jpg	10.0	7.27	3.00	3.00	10.0	5.39	4.64	4.64	10.0	4.52	3.13	3.13	1
Adachi_20170921105037.jpg	7.81	6.77	5.28	5.28	10.0	6.62	5.18	5.18	5.51	5.36	4.74	4.74	2
Adachi_20170921111658.jpg	5.72	4.43	3.65	3.65	10.0	5.66	2.42	3.24	10.0	4.93	4.30	4.30	2
Adachi_20170921143215.jpg	10.0	4.39	3.99	3.99	8.89	5.28	2.92	2.92	10.0	6.38	6.38	6.38	1
Adachi_20170921145517.jpg	10.0	4.17	2.90	2.90	5.39	2.73	2.19	2.19	5.86	4.76	3.73	3.73	1
Adachi_20170921155217.jpg	7.02	3.72	2.46	2.46	10.0	9.90	3.95	3.95	5.37	4.34	4.34	4.34	1
mean	8.48	4.80	3.56	3.54	8.48	5.68	3.89	3.86	7.59	4.73	4.18	4.17	
Standard deviation	1.93	1.85	1.07	1.08	1.79	1.45	1.08	1.08	1.93	0.71	0.75	0.76	