2023

Deep Learning Framework For Intelligent Pavement Condition Rating: A direct classification approach for regional and local roads

Waqar Shahid Qureshi
Technological University Dublin, waqar.qureshi@tudublin.ie

David Power
Pavement Management Services Private Limited, Athenry, County Galway, Ireland

Ihsan Ullah
Insight SFI Research Center for Data Analytics, University of Galway, Galway, Ireland

See next page for additional authors

Recommended Citation

Qureshi, Waqar Shahid; Power, David; Ullah, Ihsan; Mulry, Brian; Feighan, Kieran; McKeever, Susan; and O'Sullivan, Dympna, "Deep Learning Framework For Intelligent Pavement Condition Rating: A direct classification approach for regional and local roads" (2023). *Articles*. 191.
https://arrow.tudublin.ie/creaart/191

Funder: Dr. Waqar S. Qureshi has received funding under the Marrie Currie Career-Fit plus postdoctoral fellowship program from Enterprise Ireland Grant No. M.F. 2021 0273, partially funded by the European Union's Horizon 2020 Research and Innovation Program under the Marie Sokolowski-Curie Co-funding of regional, national, and international programs Grant agreement No: 847402.

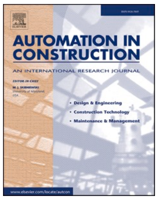
Authors

Waqar Shahid Qureshi, David Power, Ihsan Ullah, Brian Mulry, Kieran Feighan, Susan McKeever, and Dympna O'Sullivan

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon



Deep learning framework for intelligent pavement condition rating: A direct classification approach for regional and local roads

Waqar S. Qureshi [a,*], David Power [b], Ihsan Ullah [c], Brian Mulry [b], Kieran Feighan [b], Susan McKeever [a], Dympna O'Sullivan [a]

[a] *Centre for Sustainable Digital Technologies, School of Computer Science, Technological University Dublin, Dublin 7, Ireland*
[b] *Pavement Management Services Private Limited, Athenry, County Galway, Ireland*
[c] *Insight SFI Research Center for Data Analytics, University of Galway, Galway, Ireland*

ARTICLE INFO

Keywords:
Image-classification
Image-segmentation
Deep-learning-based networks
Pavement condition rating index

ABSTRACT

Transport authorities rely on pavement characteristics to determine a pavement condition rating index. However, manually computing ratings can be a tedious, subjective, time-consuming, and training-intensive process. This paper presents a deep-learning framework for automatically rating the condition of rural road pavements using digital images captured from a dashboard-mounted camera. The framework includes pavement segmentation, data cleaning, image cropping and resizing, and pavement condition rating classification. A dataset of images, captured from diverse roads in Ireland and rated by two expert raters using the pavement surface condition index (PSCI) scale, was created. Deep-learning models were developed to perform pavement segmentation and condition rating classification. The automated PSCI rating achieved an average Cohen Kappa score and F1-score of 0.9 and 0.85, respectively, across 1–10 rating classes on an independent test set. The incorporation of unique image augmentation during training enabled the models to exhibit increased robustness against variations in background and clutter.

1. Introduction

Transport departments regularly inspect pavement (road) surfaces to assess the surface condition. Pavement deterioration is primarily due to traffic, weather, and sunlight. Pavement or road surfaces can be categorized into four general classes, i.e., asphalt, concrete, gravel, and brick and block [1]. Asphalt, also known as flexible pavement, is widely used to construct national, regional, or local roads across the road network and has different sub-categories depending on its construction. Over 90% of the total European road network has an asphalt surface.

Maintenance and improvement of pavements is expensive. For example, Ireland's government spent 850 million euros in 2021 to improve and maintain local, regional, and national primary and secondary roads [2]. In Ireland, there are 5413 km of national highways (primary, secondary, and motorways), 13,124 km of regional roads, and 81,300 km of local roads adding to a total of 99,830 km of road network. The manual visual rating of pavements conducted for regional and local roads is subjective and expensive, requiring time and cognitive skills built through extensive training and experience. Automated methods for pavement assessment need to be faster, more reliable, and economical.

Pavement surface distresses in different geographical regions can be divided into six groups, i.e., cracks, surface openings, surface deformation, surface defects, joint deficiencies, and miscellaneous distresses [3]. Distresses apparent on the surface may be simply due to wear and tear or may indicate a fault in the construction. Distresses can differ in scale and appearance between rural and urban regions, depending on the surface type, the severity (low, medium, high) of the underlying problem, and other environmental conditions. Distresses can generally be detected through visual inspection (standard practice) of pavement surfaces, and their quantity and severity can be recorded using manual measurement tools [4].

Local authorities use pavement condition rating indices, incorporating all or a subset of pavement characteristics, to rate the pavement condition. These condition rating systems vary from country to country (or within a state in the USA), considering local variations, the characteristics of the pavements, the environmental condition, and the

* Corresponding author.
E-mail addresses: waqar.qureshi@tudublin.ie (W.S. Qureshi), davidpower@pms.ie (D. Power), ihsan.ullah@insight-centre.org (I. Ullah), brianmulry@pms.ie (B. Mulry), kfeighan@pms.ie (K. Feighan), susan.mckeever@tudublin.ie (S. McKeever), dympna.osullivan@tudublin.ie (D. O'Sullivan).

https://doi.org/10.1016/j.autcon.2023.104945
Received 18 January 2023; Received in revised form 17 May 2023; Accepted 19 May 2023
Available online 12 June 2023

economic conditions [3]. Pavement condition is assessed through visual surveying and usually consists of three steps: 1) pavement data collection, 2) distress identification and quantification, and 3) assigning a pavement rating index to a stretch of pavement using a standard rating scale (e.g., pavement surface evaluation rating - PASER [5]) that is typically localized to a specific geographical region [6].

In step 1, data for pavement condition assessment is usually acquired from 2D or 3D sensors mounted on a vehicle with a computer and GPS (global positioning system). Two configurations are usually used for sensor placement - externally mounted 3D sensors with a 'top-view', or internally mounted 2D cameras on the front of a dashboard giving a 'frontal-view.' The top view gives a higher ground sampling distance but covers less area per image than wide-view images. Vehicles with external 3D sensors are more expensive to operate and maintain than vehicles with an internal high-resolution camera with a frontal view [3]. In step 2, engineers manually visually inspect the collected data to identify and quantify various pavement distresses. The quantity and severity of distresses is used to compute and apply a pavement rating index for a given stretch of road in step 3. Images are captured every approximately x meters, and in practice, ratings are given to continuous stretches of roads with a similar condition, with 'y' meters (where, $y_{meters} = x_{meters}*N_{images}$) being the minimum length to have its' distinct rating. When rating images captured by a video camera, a rating is given by a data analyst viewing images on a computer offline. The rating expert assigns a rating to the first 'x' meters and adjusts the rating as the pavement condition changes.

There are several standards for visual surface assessment, including Pavement Surface Evaluation Rating (PASER) [5], Pavement Condition Index (PCI) [7,8], Pavement Surface Condition Index (PSCI) [9], and the Road Condition Indicator (RCI) [10,11]. PASER [5] is a direct rating on a scale of 10–1 (9–10 is excellent condition, while 2–1 is extremely poor). The ASTM (American Society for Testing and Materials) standard for pavement is PCI, a rating on a scale of 100–0 (100–85 is a good condition, while 10–0 completely deteriorated [7]. The Irish PSCI [9,12,13] rating is on a scale from 1 to 10, like PASER, where index-1 is the lowest (surface wholly worn out or failed), and index-10 (no distress, new pavement) is the highest. Standard ratings differ in scale granularity, pavement characteristics for evaluating the conditions localized to the geographic region and the formula to estimate a value on the rating scale.

The PSCI rating system (see Table 1) is based solely on visual pavement distresses. The impact of surface-related distresses, structural-related distresses, and other defects that affect the overall rating system is identified. Table 1 shows the primary rating indicators, the identified distresses, secondary rating indicators, the quantification measure of surface and structure quality, and the visual colour bands used for treatment measures.

In this paper, we developed a deep learning framework for direct automated pavement rating which aims to ensure more consistent and accurate pavement condition ratings and as well as to reduce the overall

Table 1
PSCI rating system and treatment measures for asphalt pavement [9].

PSCI rating	Primary rating indicators	Secondary rating indicators	Treatment measures	Surface	Structure
10	**No Visible Defects.**	Road surface in perfect condition.	Routine Maintenance	Excellent	Good
9	**Minor Surface Defects.** Ravelling or Bleeding <10%	Road surface in very good condition.		Very Good	
8	**Moderate Surface Defects.** Ravelling or Bleeding 10% to 30%.	Little or No Other defects.	Resealing & Restoration of Skid Resistance	Fair	
7	**Extensive Surface Defects.** Ravelling or Bleeding >30%.	Little or No Other defects. Old surface with aged appearance.		Poor	
6	**Moderate Other Pavement Defects.** Other Cracking <20%. Patching generally in Good condition. Surface Distortion requiring some reduction in speed.	Surface defects may be present. No structural distress[3].	Surface Restoration	Fair	Fair
5	**Significant Other Pavement Defects.** Other Cracking >20%. Patching in Fair condition. Surface Distortion requiring reduction in speed.	Surface defects may be present. Very localized structural distress (< 5 m^2 or a few isolated potholes).	Carry out localized repairs and treat with surface treatment or thin overlay.	Poor	
4	**Structural Distress Present.** Rutting, Alligator Cracking or Poor Patching for 5% to 25%. Short lengths of Edge Breakup/ Cracking. Frequent Potholes.	Other defects may be present.	Structural Overlay	Poor Overall	Poor Overall
3	**Significant Areas of Structural Distress.** Rutting, Alligator Cracking or Poor Patching for 25% to 50%. Continuous lengths with Edge Breakup/ Cracking. More frequent Potholes.	Other defects may be present.	Required to strengthen road. Localized patching and repairs required prior to overlay.		
2	**Large Areas of Structural Distress.** Rutting, Alligator Cracking or Very Poor Patching for >50%. Severe Rutting (> 75 mm). Extensive Very Poor Patching. Many Potholes.	Very difficult to drive on.	Road Reconstruction -	Very Poor Overall	
1	**Extensive Structural Distress.** Road Disintegration of surface. Pavement Failure. Many large and deep Potholes. Extensive Failed Patching.	Severe Deterioration. Virtually undriveable.	Needs full depth reconstruction with extensive base repair.	Failed Overall	

time required for pavement rating. We propose a deep learning approach using image segmentation and classification methods.

1.1. Related work

Researchers have proposed several methods for automating visual surface condition assessment based on computer vision, machine learning, and, more recently, deep learning [14–17]. Researchers have recently reviewed various automated pavement distress detection and data acquisition, including 1D sensors, 2D sensors, and 3D sensors [18–24]. Much of the literature focuses on automating step 2 – automatic distress identification and quantification - while there is less emphasis on automating step 3 – directly computing a pavement index rating from image data.

1.1.1. Digital data collection and pavement image datasets for condition assessment

Digital data collection is essential to automating pavement distress identification, quantification, or direct condition assessment. A guide on data collection, including visual data for pavement quality management, is presented in [25]. The guideline presents standard procedures and practices to obtain data for pavement quality assessment and management. For visual pavement condition assessment through images, either of two views is recommended, i.e., a front-mounted camera placed orthogonal to pavement surface normal, or a back-mounted camera placed inline to pavement surface normal. However, vehicles with external 3D sensors are more expensive to operate and maintain than vehicles with an internal high-resolution camera with a frontal view [3].

In [21], the authors list contributions to existing publicly available pavement image datasets for distress detection. These very limited datasets can be categorized based on the view angles (top-view, wide-view, hand-held), and imaging technologies (3D or intensity) mainly focused only on a subset of distress types (different crack types, potholes, and patches) found locally in the geographical regions (USA, China, India, Japan, Czech Republic, Brazil, Italy, and Mexico). Two frontal view datasets focus on pavement rating; the first is the Paris-Saclay and the second is the Road Quality Dataset (R.Q.) [26]. The Paris-Saclay dataset [27] is annotated for pavement condition rating for a stretch of a road based on PASER for New York roads. The frontal-view images are extracted from Google Maps API, while the ground truth annotation for each stretch is extracted from the pavement condition rating of New York [28]. The ground truth annotation contains the street index, the number of images in the street, the PASER rating for each street segment, and the course rating of good, fair, and poor for each street segment. A similar image dataset can be extracted from Google images for Oakland, USA, while the street segment pavement rating based on PCI can be generated from the database available [29]. R.Q. Dataset [26] is a manually annotated frontal-view image for pavement condition index ratings based on six different condition ratings for the Czech Republic. The pavement condition rating criteria are defined in [26], while the images are obtained using Google Maps API. The image dataset annotated for pavement rating indices is also limited and does not cover the full range of standard visual rating scales, i.e., PASER and PSCI. Over the years, researchers have made available datasets for benchmarking automated distress detection systems, mainly covering different types of cracking and potholes [30]. Only a few focus on other distresses detection or visual pavement condition rating classification.

1.1.2. Automated distress detection and identification

Deep learning architectures have recently been applied to pavement condition detection and classification [31–48]. These methods can be segregated into pavement condition rating through image classification, pavement distress detection using object detection, and semantic segmentation approaches for pavement cracking. Researchers in [34,49] have used aerial images through drones as input and presented a convolutional neural network architecture for automated pavement distress detection (mainly cracks) and evaluation, respectively. In [48], an automated smartphone-based application is proposed to detect potholes and cracks. An accelerometer, global positioning system (GPS) sensor, and compass are used to record the location of the potholes. Recall, precision and accuracy are reported for eight distresses, with the lowest recall recorded as 5% for lateral linear cracks, 65% for alligator cracking, and the highest for crosswalk blur and white line blur at 95%. Authors in [22,41,50] also used a smartphone mounted on the dashboard of a vehicle to capture images from multiple countries and develop distress detectors, based on CNN, for alligator cracks, longitudinal cracks, transverse cracks, and potholes. The distress objects are similar to [48], and the measures reported for the three countries are F1-score and mean average precision. Pavements in different countries have different F1-score with a maximum F1-score of 52% for alligator cracking and a minimum F1-score of 29% for linear transverse cracking for Japanese roads. In [51], a CNN-based crack segmentation method consists of a novel architecture of five layers; the input layer is a line feature detector filter, followed by two convolutional layers and two fully connected layers to segment crack pixels in the 3D images of asphalt surfaces. The evaluation reported precision, recall, and F1-score with an F1-score of 88%. This method is specifically for 3D data from the PaveVision3D laser system, which is mounted on a video van, viewing an orthogonal top view of the road.

In [52] authors present the first CNN-based ravelling detection by training macro texture features obtained from the 3D images from PaveVision3D [53]. In [54] authors propose an automatic patch detection system using object detection techniques with state-of-the-art models (Faster RCNN and SSD MobileNet-V2). The results show successful detection of patches in LCMS images, suggesting integration with existing systems.

In general, most classification-based approaches focus on identifying types of distress in an image patch of higher-resolution images. Localized distresses are investigated, i.e., potholes and cracks. Patch-based classification and identification of distress instances are helpful for localized distresses; the technique is suitable for images that capture the top view of the road. The deep-learning-based segmentation algorithms perform well when the test data is similar to the training images (i.e., from the same device); however, the performance degrades when the multiple training datasets are combined, or the test dataset is from a different capturing device and region. The performance of the object detector-based distress detection deteriorates for multiple distress detection compared to detectors that detect one or two distresses [3]. Recall or accuracy for detecting cracks (linear or edge) using a frontal view image is less when object detection networks such as Yolo [55] are used compared to top-view images. Methods that focus on automated distress detection and identification are not directly able to estimate the pavement condition rating on a standard rating scale.

Most of the work in automated image-analysis-based pavement condition assessment is focused on two primary distress types, i.e., distinct types of cracks and potholes. For PASER (used in the U.S. and other regions) and PSCI (used in Ireland), the ratings 10–7 are decided based on the amount of ravelling and bleeding alone. Very few experiments can be seen in the literature on ravelling or bleeding distresses (see [56,57]), which are forms of surface defects and contribute towards a unified pavement surface rating. Other surface distress such as patching, utility patches, and utility cover is also seldom considered (see [54]).

1.1.3. Direct pavement rating

The primary purpose of distress detection and identification is to evaluate the pavement condition using a standardized scale. Distresses must first be identified and then the number of distinct distresses and their severity must be considered over a given stretch of pavement to compute a rating. Most research focuses on distress identification (see Section 1.1.2) but falls short of automatically computing a direct pavement rating for a stretch of pavement. One approach to computing direct ratings is described in [45]. The authors present a hybrid model of an object detector and semantic segmentation for classifying and quantifying distress severity on pavements and predicted PASER indices for each patch. The images are collected from Google Street View maps, 70-degree wide-angle views, and 90-degree birds-eye view images. Wide-view photos are used for crack and pothole detection, and top-view images to quantify crack severity. The results from the hybrid model are then fed to a linear and weighted regressor for predicting PASER indices. A YOLO model was trained to detect distresses (cracks and potholes) and a U-Net (based on a fully convolutional layer) model to classify crack severity. The results from the two models are then combined to find the crack density per pavement defect. The results are then fed to a linear and a weightage regressor to label each image a PASER index. The predicted PASER model fits with an R^2 of 0.9382 or test data with a root mean square error of 10.45. One of the limitations of this research is the use of Google API images that are quite old. In this system, only two distresses are used to compute the rating (cracks and potholes); however, in most practical scenarios, cracks, potholes, patches, ravelling, and bleeding also need to be considered, requiring transfer learning for adding localized distresses to the algorithm.

In [39], the authors present an image classification approach to surface rating where they used a three-rating index - good, regular, and bad. The dataset used for the experiments is RTK [46], caRINE [58], and KITTI [59]. It classifies roads into three different types (unpaved, paved, asphalt) and three different ratings (good, regular, bad). The surface type accuracy is reported to be 98% for three types. The classification accuracy for the three asphalt quality types is 98% for good and 96% for bad. The precision of classifying the good class is 86.7% while classifying the bad asphalt class is 81%. The number of rating indices is limited to three – good, bad, and regular and judging on a scale of 3 levels is not very useful in real life. Maintenance discussions are based on the overall rating and the individual distresses that lead to that rating. As such further experiments are required to increase the number of classes useful for visual standards such as PASER or PSCI.

In [60], the authors used pixel segmentation using a semantic segmentation CNN-based model from [61] to extract roads, marks, and background pixels. They analyzed the state-of-the-art EfficientNet V2 [62] image classification approach for automating PSCI ratings. Each image in the training and test set has a 'segmented' pavement image, an 'augmented' image, and an 'original' image. Image height is cropped 250 pixels from the top and 50 pixels from the bottom to remove the sky and pavement pixels further away from the camera and pavement pixels too close to the camera. The "Augmented" images are computed by combining the pavement segmented intensity image, the pavement plus mark pixel intensity image, and the original intensity image. They used a combination of these images to evaluate the performance of their classifier. For a 10-class classification based on PSCI, the best model achieved an F1-score of 0.57, while an F1-score of 0.73 was achieved for five-class classification after combining adjacent classes.

On the commercial side, a few companies in the U.S and Japan do provide automated solutions for pavement condition ratings. RoadBotics [63], working locally for U.S roads, use a limited version of PASER [5], i.e., they rate pavement condition from 1 to 5. An automated rating system from Ricoh [64] estimates the amount and location of cracks on 50 cm × 50 cm patches and has adopted its rating system for Japanese roads based on PCI.

The current literature on pavement condition assessment is limited in its scope. Existing methods focus on specific types of distress present in a particular region, use an orthogonal view of the pavement, or are focused solely on distress detection and identification. Little work has been done on direct pavement condition rating classification using a low-cost camera mounted on the front of a vehicle, which is a low-cost approach for pavement condition assessment for regional and local roads. Experts typically use standardized rating scales, such as PCI, PASER, or PSCI, to provide an objective assessment of pavement condition rating or severity classification. These non-linear scales consider various factors, including the type, extent, and severity of pavement distress, and provide an overall rating based on visual distress, including surface-related and structural-related issues. Table 1 outlines the visual distress rating criteria for PSCI. To maintain quality control, it is common for multiple experts to rate an image, and inter-rater reliability is calculated using the Cohen's kappa score. This statistical measure considers the agreement between two or more raters beyond chance agreement and ranges from -1 to 1, with values closer to 1 indicating stronger agreement between raters. While automated pavement distress detection and quantification methods exist, they are not yet accurate or consistent and may require extensive labelled data, person-hours, and complex algorithms to develop a non-linear model based on deep neural networks. To address this, we hypothesize that a well-designed training set capturing different distributions of visual distress and severity with reliable expert applied labels can enable the creation of a direct DL-based classifier.

Direct classification offers a comprehensive approach to evaluating pavement condition by directly identifying and classifying various types of distresses present in pavement images. Unlike quantified calculation approaches, direct classification considers the overall condition of the pavement and can capture subtle variations and complex patterns that may be challenging to accurately quantify using predefined metrics. This flexibility allows direct classification to adapt to different types of distresses and variations in pavement conditions without relying on specific predefined rules or thresholds.

However, it is important to note that direct classification requires labelled pavement images with accurate annotations for each rating scale. To ensure the models can generalize and accurately classify pavement conditions across different scenarios, they need to be trained on a diverse and representative dataset. This dataset should encompass various pavement conditions and include a wide range of distress types and severities.

We propose a deep neural network framework for direct pavement condition rating specifically designed for flexible asphalt regional and local roads. Unlike the traditional approach of distress detection, quantification, and rating, which can be computationally complex and time-consuming, our framework employs a direct classification methodology, simplifying the assessment process. The framework leverages the Pavement Surface Condition Index (PSCI) rating (see Table 1), which establishes a relationship between primary and secondary distresses and their objective measures, treatment measures, and defect segregation as required by local authorities. Within our framework, we have developed deep learning-based models for two key components: pavement segmentation and condition rating classification blocks. These models are integrated into the proposed architecture to enable accurate pavement assessment. To achieve this, we trained and evaluated various state-of-the-art deep learning architectures, including transformer-type models such as DeepLabV3 and SegFormer, as well as convolutional neural network (CNN) models like ConvNeXt, ResNet50, and Swin-V2.

In the first stage, the deep learning model is developed to extract pavement pixels from the input images, enabling precise pavement segmentation. This pixel-level segmentation is crucial for isolating the pavement surface from the background and other objects, ensuring accurate condition assessment. In the second stage, the deep learning model is developed to classify the pavement condition based on the PSCI rating system. By examining the segmented pavement regions, the models assign the corresponding PSCI condition rating to each image, providing an overall assessment of the pavement's condition. To enhance the robustness to different background changes and clutter we incorporated unique image augmentation technique during the training process.

By combining the direct classification methodology, deep learning-based models, and innovative image augmentation techniques, our framework offers a comprehensive and efficient solution for pavement condition rating. The integration of the PSCI rating system ensures that the assessment aligns with the objective measures, treatment measures, and defect segregation required by local authorities.

Our method incorporates a comprehensive visual inspection of the pavement surface, encompassing various distresses including cracks, potholes, surface deterioration, and other visible signs of distress. By considering this wider range of defects, our method provides a more holistic assessment of pavement condition, allowing for better prioritization of maintenance and rehabilitation efforts. Our framework and augmentation techniques are generalizable to other visual inspection-based rating system such as PASER and implicitly account for broader range of distresses, however would require specific training data to finetune the model.

We collected a large dataset of publicly available images captured from various regional and local roads in Ireland. The dataset is rated using the PSCI rating scale from 1 to 10 by two expert raters who have years of experience in pavement condition rating using the standard rating scale. Our framework comprises a number of pipelined blocks - pavement-segmentation, data (image) cleaning, image cropping and resizing, and image classification. To test our framework, we developed a dataset of 7453 images captured from different regional and local roads in Ireland, which are rated on a PSCI rating scale by a team of experts. The road segmentation model was trained on 3349 images, and the PSCI rating classification model on 4581 images. In the sections to follow, we outline the methodology, including dataset, deep learning frameworks, training parameters, and evaluation criteria. In Section 3 we present the results of the framework, and we conclude with a discussion in Section 4.

2. Material and methods

The input to our pipelined framework for direct pavement rating classification are images of flexible asphalt pavements of regional and local roads from urban and rural environments across Ireland. The images are acquired using a camera mounted on the front dashboard of a video van (see Fig. 1(a)). The camera is attached to a server for recording images. A remote laptop accesses the server over the network to label each image stretch of the pavement—the server linked to the camera capture image every five meters.

The image deep learning framework consists of number of blocks – semantic (object) segmentation, image processing and, PSCI rating image classification, and extract and blur block (see Fig. 2). The semantic segmentation and classification blocks are deep-learning-based models. Pavement segmentation infers a segmented pavement image using semantic segmentation. Image processing eliminates poorly segmented image pairs and crops and resizes the image before feeding it to the classifier. The classification step infers a PSCI rating for each image. The extraction and blur blocks generate the final output image after blurring the human faces and vehicles by using the mask generated by the segmentation block.

In this text to follow, we first explain the image labelling, then the dataset created for training and testing, and finally explain each block of the pipelined framework.

2.1. Dataset and labelling

The size of the image captured by the camera is 720 × 576, with three channels (red, green, and blue). We created two labelled image datasets extracted from video frames captured. 1) The first dataset (IrishRoadSurvey-1) is used to develop and test our semantic segmentation model that can classify pixels into seven classes (i.e., background, human, pole, road, traffic light, traffic sign, and vehicles). 2) The second dataset (IrishRoadSurvey-2) is used to develop and test our model for image classification that can assign pavement surface condition indices (PSCI 1–10) to each pavement surface in the image.

2.1.1. Training and test set for image segmentation

IrishRoadSurvey-1 consists of five hundred colour images (374 training, 180 test) of different flexible pavement surfaces of Irish roads (urban, local, motorway, national primary, national secondary, and regional). Of these, 42% were national primary, secondary, and motorways, and 58% were local and regional roads. All pixels in both training and test images were labelled into seven different classes. In addition to IrishRoadSurvey-1, we customized the original Cityscapes [65] dataset containing 2975 training and 500 test images that has previously been labelled for 19 different classes, to train the image segmentation deep learning model. We relabelled the Cityscapes dataset into seven classes, cropped it, and resized it to a lower resolution (720 × 567). By combining the IrishRoadSurvey-1 and customized Cityscapes [65] dataset, there were 3349 training images and 680 test images for developing and evaluating a semantic segmentation model.

2.1.2. Training and test sets for image classification

The size of the available dataset is relatively small; therefore, we followed 70% training and 30% test ratio for all our experiments; the images for both sets were randomly chosen at the start. IrishRoadSurvey-2 initially consisted of 7453 colour images (5024 training, 2429 test) of flexible pavement surfaces classified into ten condition rating indices from the PSCI rating scale. The images are of the same size as IrishRoadSurvey-1. The images are not frames from a continuous video; they have been selected from different regional and local flexible asphalt pavements across Ireland. We removed image

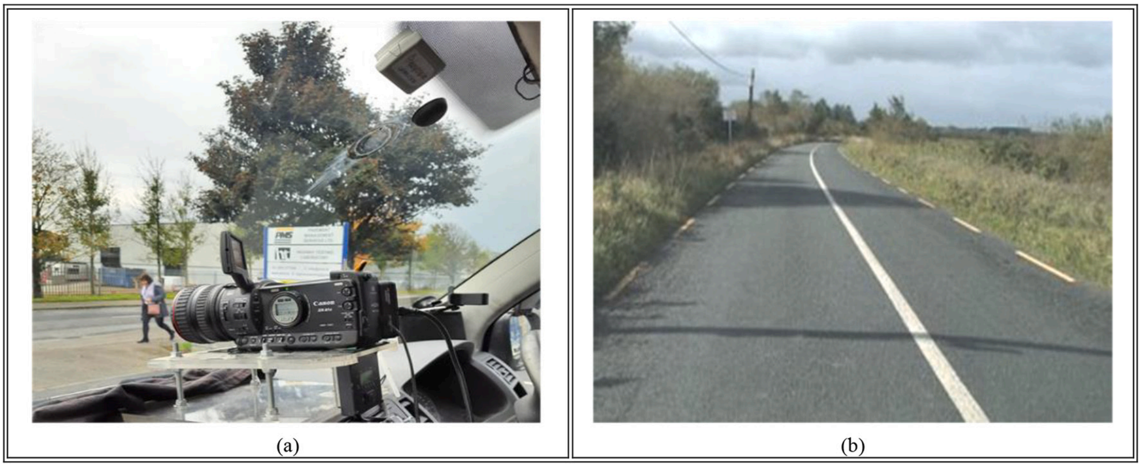


Fig. 1. Typical camera position, captured image, and the image after different pre-processing steps. (a) is the picture of a typical camera position mounted on the video van. (b) the output of the camera with a rating of 10.

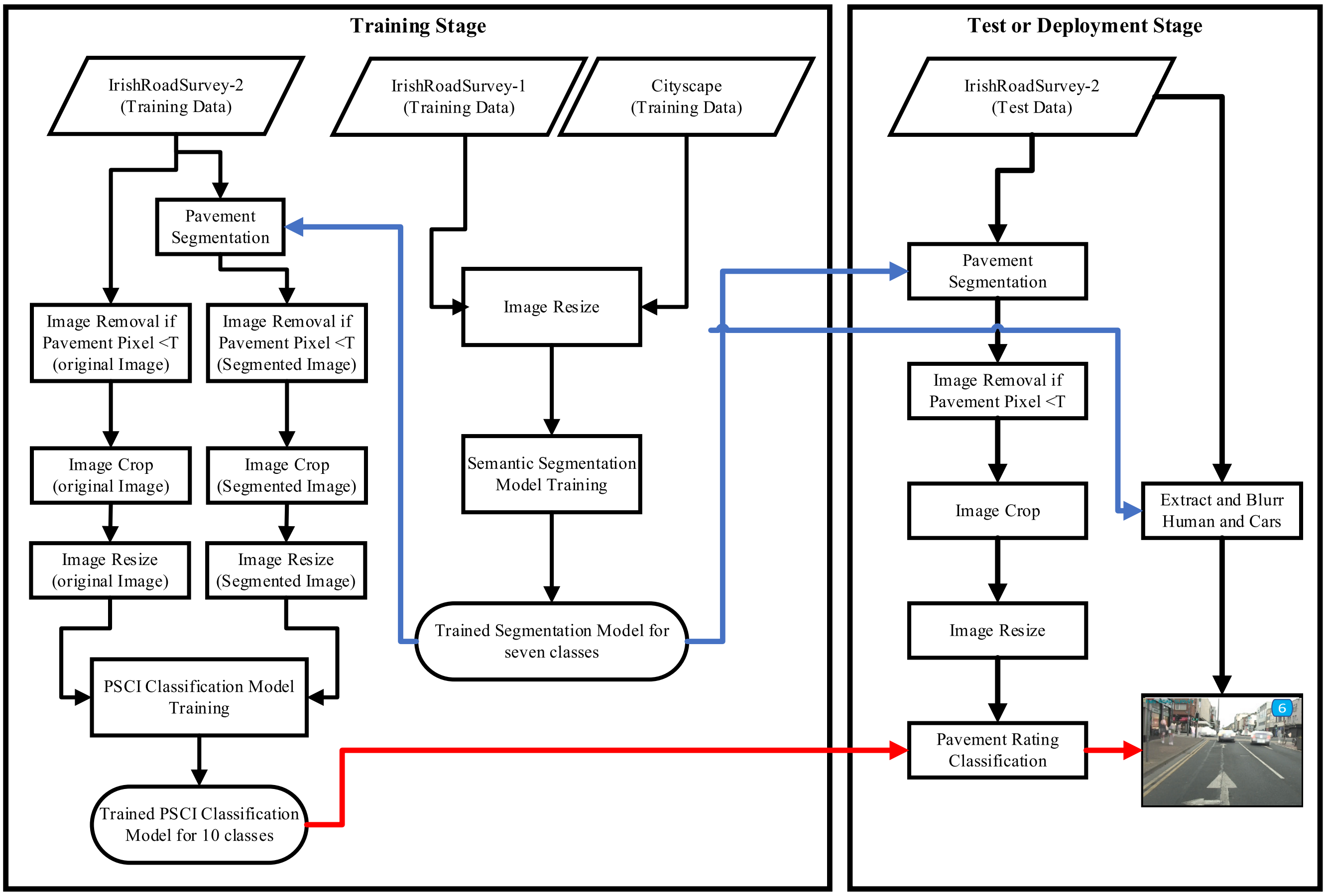

Training Stage
IrishRoadSurvey-2 (Training Data)
IrishRoadSurvey-1 (Training Data)
Cityscape (Training Data)
Pavement Segmentation
Image Resize
Image Removal if Pavement Pixel <T (original Image)
Image Removal if Pavement Pixel <T (Segmented Image)
Semantic Segmentation Model Training
Image Crop (original Image)
Image Crop (Segmented Image)
Image Resize (original Image)
Image Resize (Segmented Image)
Trained Segmentation Model for seven classes
PSCI Classification Model Training
Trained PSCI Classification Model for 10 classes
Test or Deployment Stage
IrishRoadSurvey-2 (Test Data)
Pavement Segmentation
Image Removal if Pavement Pixel <T
Image Crop
Image Resize
Pavement Rating Classification
Extract and Blurr Human and Cars
6

Fig. 2. Block diagram for Deep learning framework for an intelligent pavement condition rating direct classification.

frames with motion blur, images with insufficient lighting due to shadow from trees, and images with focus blur. The simple elimination technique (see Section 2.2.2 for details) removed around 613 images that were initially part of the IrishRoadSurvey-2 dataset. The dataset after removal of images consisted of 6855 colour images (4581 training, 2274 test). Our dataset is an extension of the PSCI dataset developed by Qureshi et.al [60] and has 56% more images than the original dataset.

Table 2 shows the total number of images in the Training and Test IrishRoadSurvey-2 dataset before and after elimination. The removed images had <25% extracted pavement pixels in the image (see image processing explanation in Section 2.2.2). As the pavement pixels were extracted using the semantic segmentation model, we found two primary reasons for poor segmentation in the eliminated images: firstly, fewer actual pavement pixels due to the camera view and zoom; secondly, fewer segmented pavement pixels due to poor performance of the segmentation algorithm on those images due to lighting or other

Table 2
No. of images available for segmentation and PSCI classification dataset and their categorisation.

Segmentation dataset						PSCI classification dataset					
S·No	Folder	Images	S·No	Folder	Images	S·No	PSCI	unclean		clean	
1	Aachen	174	1	Frankfurt	267			Train	validate	Train	validate
2	Bochum	96	2	Lindau	59	1	1	487	209	301	137
3	Bremen	316	3	Munster	174	2	2	372	160	334	135
4	Cologne	154	4	IrishRoadSurvey-1	180	3	3	394	169	384	168
5	Dardmstad	85	5	Total	680	4	4	406	175	376	168
6	Dusseldorf	221				5	5	415	170	370	153
7	Erfurt	109	Sub category of IrishRoadSurvey-1			6	6	682	571	674	562
8	Hamburg	248	S·No	Category	Images	7	7	608	261	512	244
9	Hanover	196	1	Urban	80	8	8	654	281	644	281
10	Jena	119	2	Disintegrated local	30	9	9	527	227	508	220
11	Krefeld	99	3	Local	90	10	10	479	206	478	206
12	Monchengladbach	94	4	Motorway	90	11	Sub-total	5024	2429	4581	2274
13	Strasbourg	365	5	National primary	58	12	Total	7453		6855	
14	Stuttgart	196	6	National secondary	60						
15	Tubingen	144	7	Regional	92						
16	Ulm	95	Total		500						
17	Wwimar	142									
18	Zurich	122									
19	IrishRoadSurvey-1	374									
20	Total	3349									

environmental conditions while capturing the image.

2.1.3. *Image labelling*

We used CVAT [66], to label pixels in IrishRoadSurvey-1 and labelled images in IrishRoadSurvey-2. The CVAT is open-source software for labelling and can run on a local server. It provides different labelling and annotation methods. Two labellers performed pixel-level annotation for the segmentation task; one performed the actual annotation, while the other counter-checked the labelling done by the first labeller. The labellers were directed to complete fine annotations for humans, roads, backgrounds, and vehicles while keeping course annotations for traffic lights, traffic signs, and poles.

The images in IrishRoadSurvey-2 were annotated offline using a PSCI [6] scale of 1–10 by two data analysts. Table 3 show the difference in the ratings between two data analysts while classifying RoadSurveyDataset-2. It shows a high rating agreement between the raters about the labels applied to the images used to develop the models with a weighted kappa score of 0.97. As described in Table 3, the PSCI rate was determined by two labellers (trained and experienced) who reviewed each sample independently. If there was no agreement (i.e., no exact same rating value provided by both) between the two labellers, a third expert data analyst was consulted to make the final determination. Regarding the choice of using the PSCI rate labelled by expert data analyst-1 as the ground truth, this decision was made based on our belief that expert data analyst-1 had the most experience and knowledge in this area, and therefore their labels were deemed to be the most accurate and reliable. However, we acknowledge that this decision may have introduced some bias into our dataset. We ensured that all labellers were highly trained and experienced in the task of rating PSCI. Nonetheless, in real-world analysis, such limitations are often managed by providing a confidence level of the rating assigned by the analyst and by implementing quality control loops to ensure accuracy and reliability.

For our experiments, the images are divided into classes 1–10 according to the PSCI ratings. Fig. 3 shows the sample images from the IrishRoadSurvey-2 labelled by the expert PSCI data analyst for PSCI ratings from 1 to 10. The original images were used for annotation. The classes are imbalanced and representative of a real-world scenario for pavement conditions for regional and local roads. Fig. 3 can be used along with Table 1 to understand the sample of different type of pavement condition in each rating index. For example, Image rated as 3 have more severe alligator cracking than image rated as 4. While image rated 5 has visible linear cracks; image rated 6 has clean utility patches. Images 7–9 are classified based only on the amount of bleeding and ravelling present, and do not show any other sign of distress; images rated as 10 do not have any visual distress.

In the IrishRoadSurvey-2 dataset, for each rating scale, there is a disagreement between the labellers for a few images; this disagreement is usually between adjacent rating classes. We observe that manual rating can have some randomness, especially for multiclass problems. However, we took several steps to minimize this randomness (see Table 3). The raters are working professionals and have years of experience of rating Irish roads for pavement condition ratings for local counties. Firstly, we provided raters with clear instructions on how to rate the pavement images. Secondly, we trained the raters on a subset of the dataset to ensure they were consistent in their ratings. Finally, we provided the raters with the standard PSCI manual to help clear definition for each rating category to help them make consistent and accurate ratings for each image.

2.2. *Pipelined framework blocks*

In the following text, we explain each of the individual pipeline blocks and our choice of the internal architecture of the framework shown in Fig. 2.

2.2.1. *Semantic segmentation block*

We observed in an earlier attempt [60] that the intel OpenVino library [67] provides a ready-to-use road segmentation model with a mean accuracy of 89.9% and an average IOU of 84.4% for four classes, mainly road (IOU = 95.5%), curbs, background, and road lane markings on Might AI [68] test images. Our experiments on local and regional Irish roads showed that the performance of OpenVino model [67] decreases significantly on deteriorated pavement surfaces [60], and the model provided by [67] does not allow transfer learning or retraining. Therefore, instead of developing a new architecture from scratch for road segmentation block, we evaluated existing deep learning segmentation architectures for developing seven class semantic segmentation models. We agree that the primary objective of the segmentation block in our framework is to extract the pavement from the background to analyse the PSCI. Heuristically, we found that reducing the problem to

Table 3
Confusion matrix showing agreement and disagreement between two raters for pavement surface condition rating. The diagonal shows the number of agreement images between the two raters.

		PSCI rating by Rater 2													
		1	2	3	4	5	6	7	8	9	10	Agree	Disagree	Total	
PSCI rating by Rater 1	1	683	12	1	0	0	0	0	0	0	0	683	13	696	PSCI Rating by Rater 1
	2	4	512	15	2	0	0	0	0	0	0	512	21	533	
	3	2	6	539	18	0	0	0	0	0	0	539	26	565	
	4	0	0	7	567	7	0	0	0	0	0	567	14	581	
	5	0	0	0	6	559	17	0	0	0	0	559	23	582	
	6	0	0	0	0	23	1218	8	0	3	1	1218	35	1253	
	7	0	0	0	0	0	3	851	13	2	0	851	18	869	
	8	0	0	0	0	0	0	7	912	15	1	912	23	935	
	9	0	0	0	0	0	0	0	20	724	10	724	30	754	
	10	0	0	0	0	0	0	0	0	11	674	674	11	685	
	Totals	689	530	562	593	589	1238	866	945	755	686	7239	214	7453	
	Disagree	6	18	23	26	30	20	15	33	31	12	Weighted Kappa			0.979
		PSCI Rating by Rater 2													

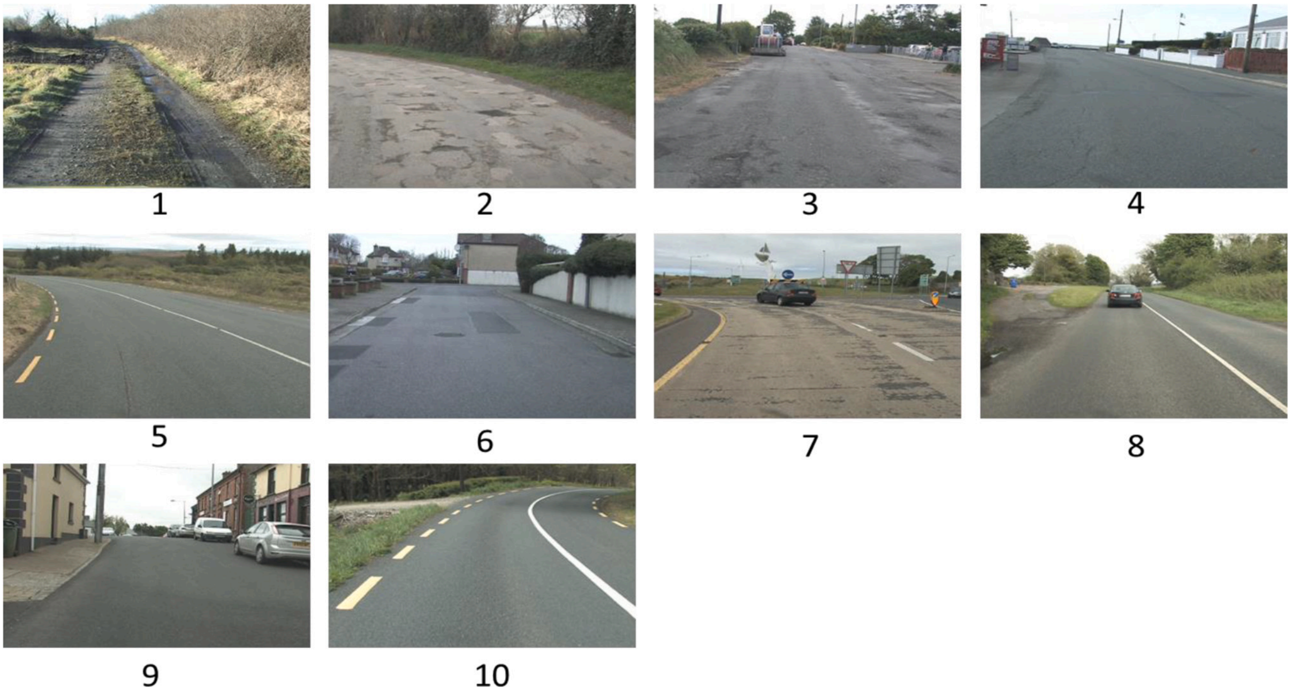


Fig. 3. Sample images from the RoadSurveyDataset-2 labelled by the expert PSCI rater for PSCI from 1 to 10.

binary classification (i.e., pavement and background), decreases the intersection of union for pavement pixels due to the imbalance of pixels between the two classes. To develop an image segmentation model, we labelled the images for the common features present in the image i.e., human, pole, road, traffic light, traffic sign, and vehicles, which has shown to improve the accuracy of pavement IOU as compared to binary classification of pixels, especially for those classes for which we have sufficient training data.

The semantic segmentation-based deep learning architecture usually consists of two sub-blocks, the encoder and the decoder. The encoder usually consists of a stem block and feature extraction block, while the decoder generates the semantic segmentation mask using high-level features. we compared Deeplab v3 - with ResNet-50 encoder, and SegFormer with Transformer encoder. Below, we briefly explain the architectures we evaluate for our segmentation task.

2.2.1.1. SegFormer. SegFormer [69] is pixel-segmentation architecture consisting of Mix Transformer encoders (MiT), a lightweight multilayer perception (MLP) decoder, and a 1 × 1 convolutional head for generating semantic segmentation masks. SegFormer initially proposed six different MiT sizes with the same architecture. We choose to use MiT-B2 with moderate memory utilization and learning parameters. The novelty that makes SegFormer different from other segmentation algorithms is the hierarchical encoder which enables it to compute the deep features, allowing features to capture information at different spatial scales. As opposed to vision transformers, ViT [7], SegFormer does not need positional encoding and increases performance when the testing resolution is different and the training resolution. The encoder hierarchically has four transformer blocks (T.F.), with a normalization layer after block. Each T.F. has an overlap patch merging followed by a self-Attention block and Mix FFN (a dense layer followed by the 3 × 3 depth-wise convolution and a Gaussian Error linear Unit). The MLP decoder enables the SegFormer to build up information from different encoder layers, combining local and global attention to a dense layer that produces a semantic segmentation mask (see Fig. 2 in [69]). The encoder head we used was trained on ImageNet-1 K, while the decoder head was learned using fine-tuning technique on our customized dataset for the semantic segmentation model.

2.2.1.2. DeepLabV3. DeepLabV3 [70] is an image segmentation deep learning architecture composed of fully convolutional layers. Usually, in an encoder-decoder-head type semantic segmentation architecture, the image size is reduced to reduce the number of parameters and computations. Their encoder architecture uses “atrous “convolutional filters, a filter-up sampling technique applied to the original high-resolution image. This is achieved by up sampling the kernel filter and introducing zeros between the filter sample instead of interpolating. The encoder consists of a standard deep feature detector, which in our case is ResNet-50 [71] consisting of five stages of convolutional blocks. We used two separate instances of ResNet i.e., ResNet50 [71] and ResNet50b [71,72], which is ResNet architecture with 50 layers and a sampling rate of eight. Here ‘b’ stands for a blur pooling. ResNet-50 is followed by a depth-wise separable Atrous convolution and Atrous Spatial Pyramid Pooling (ASPP) decoder. The ASPP decoder consists of ASPP modules (i.e., layers of depth-wise separable convolution modules, batch normalization, ReLU) and an image pooling layer. The decoding layer helps to classify each pixel corresponding to one of the seven classes. The output from the decoder is attached to a 1 × 1 convolution layer to get the final segmentation mask. The architecture used in the current study is slightly different from than the standard version. A separate, fully connected auxiliary head is attached for better optimization of the models used in PSPnet [73].

2.2.2. Pavement segmentation and image processing

For each image, the semantic segmentation block generates a mask, which is applied to the original image to compute an image containing only pavement pixels and is called a ‘segmented image.’ They are then passed to the image processing block along with the corresponding original image. The image processing block prepares images for the PSCI classifier. In the image processing block, we first find the images with pavement pixels less than a threshold ‘T’ in the segmented image and then remove them and their corresponding original image pair. The relation is shown in the equation Eq 1 below, where p is the pavement pixels in the image ‘x,’ and y is the binary variable to decide on keeping the image set for training.

$$\text{if} \sum p_x > \mathrm{T}\ \mathrm{y}_\mathrm{x} = \mathrm{True} \tag{1}$$

where T = (720 x 576)*0.35

Then for each image set (i.e., 'segmented image' and original image), its height is cropped by 246 pixels from the top to remove the sky. Its width is cropped by 10 pixels on each size to make the resolution to 700 × 330. The image set is then resized (224 × 224 or 384 × 384) to make it ready for the classification block. Fig. 4 show different stages of cropping and resizing of image before passing to the classifier.

2.2.3. PSCI classifier block

To develop a deep learning PSCI classifier for the classification pavements on a standard rating scale (PSCI 1–10) from pavement pixels extracted from images, we used a fine-tuning approach inspired by [74–76]. A deep learning architecture for classification usually consists of a stem block, layers of feature blocks, and a classification head. A stem block down samples the input image to an appropriate feature map. The feature block extract' deep features for each image passed to the architecture. A fully connected neural network layer known as 'classification-head' is attached to the feature block for any image classification task. We developed three architectures, R, C, and S, for PSCI (1−10) classifier by integrating a classification head on top of a ResNet50 [72], ConvNeXt [77], and the transformer-based Swin-V2 [78] feature blocks, respectively. We then compare our model results with the EfficientNet V2 model proposed by [60]. Since such applications must run on limited hardware resources (Embedded hardware or a tablet), our focus is to choose models with higher accuracy with limited memory resources (i.e., fewer parameters) and lesser floating-point computations and, therefore, smaller architectures. ResNet50 is the baseline, while ConvNeXt [77] and Swin-V2 [78] are state-of-the-art classifiers from the convolutional and transformer bases, respectively.

ResNet50 [71] which has 50 deep layers, is a ResNet family architecture usually considered a baseline method for comparing neural networks. It uses residual bottleneck blocks for better convergence. The stem block contains a convolutional layer consisting of a 7 × 7 kernel of 64 filters, followed by batch normalization layer, rectified linear unit (ReLU), and Max pooling to down sample four times the size of the input image. The feature block consists of four sequential layers with 3, 4, 6, and 3 bottleneck residual blocks in the first, second, third, and fourth layer, respectively. The filter channels are expanded from 64, 128,256,512, 1024, and 2048 of size 3 × 3 filters in a hierarchical way towards the head. The bottleneck residual block consists of two 1 × 1 convolutional layers and one 3 × 3 convolutional layer. The 1 × 1 convolutional layers increase the depth and reduce the parameters by down sampling and then up sampling after the 3 × 3 convolution layer in the block. A batch normalization and ReLU layer follow each convolutional layer. At the end of the residual bottleneck block, the output is added to the input of the residual block for better convergence of the optimization function. The ResNet50 backbone (stem and feature block) architecture output is a 8×8 × 2048 feature map (see Fig. 5). A linear, fully connected neural network layer is connected to the feature block to develop an image classifier 'R' model.

ConvNeXt [77] is a recent work from Facebook research from 2022 that argues the potential of CNN architectures over transformers-based architectures by redesigning the ResNet architecture family [71]. They use larger filters, change to layer normalization from batch normalization, use GELU [79] over ReLU, and separate down sampling layers. It uses a 96-channel filter of size 4 × 4 with a non-overlapping (i.e., stride 4) strategy in the stem block. It works similarly to segmenting the images into patches and applying feature mapping filters. The feature block consists of four sequential layers with 3, 3, 9, and 3 ConvNeXt blocks in the first, second, third, and fourth layer, respectively (see Fig. 6). A ConvNeXt blocks consist of a 7 × 7 convolution kernel with d-channels, where d varies from 96, 192, 384, and 768 in the four sequential layers. A bigger version of the ConvNeXt-tiny, also used in our evaluation, with similar channels but different configurations (i.e., 3,3,27, and 3) of the blocks in each sequential layer is referred to as 'ConvNeXt-small.' The convolution layer is followed by a layer normalization layer and a Multilayer perceptron (MLP - a two-layer fully connected feedforward neural network). Like the residual bottleneck, at the end of the convNeXt, the output is added to the input of the block (skip connections) for better convergence of the optimization function. The ConvNeXt backbone (stem and feature block) architecture output is a 12 × 12 × 768feature map (see Fig. 6). A linear, fully connected layer is connected to develop an image classifier 'C' models.

Swin-V2 [78] is a state-of-the-art hierarchical vision transformer-

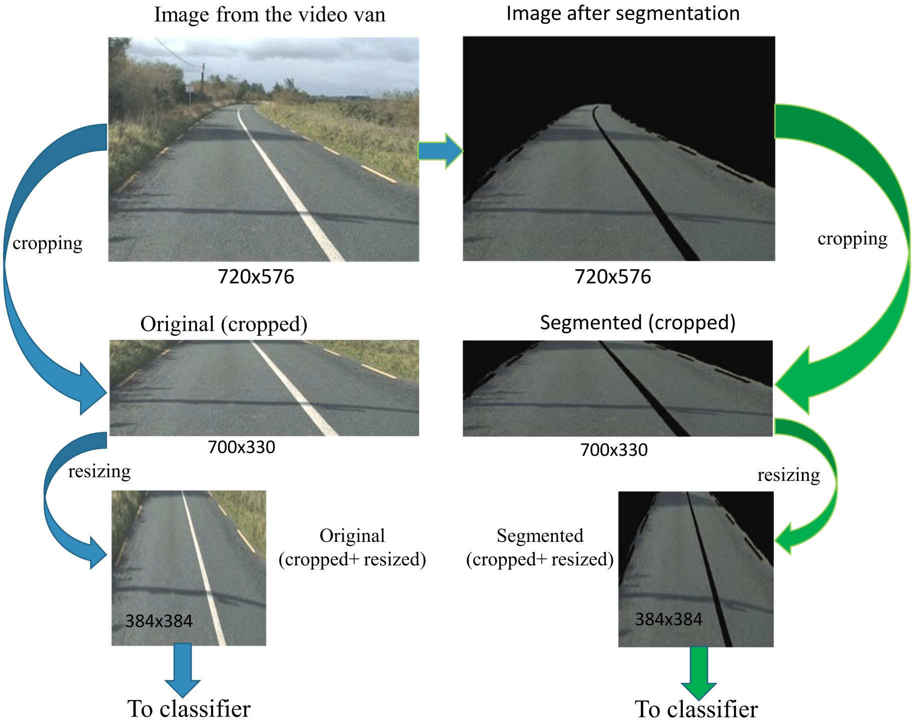


Fig. 4. Different stages of cropping and resizing before passing to the classifier.

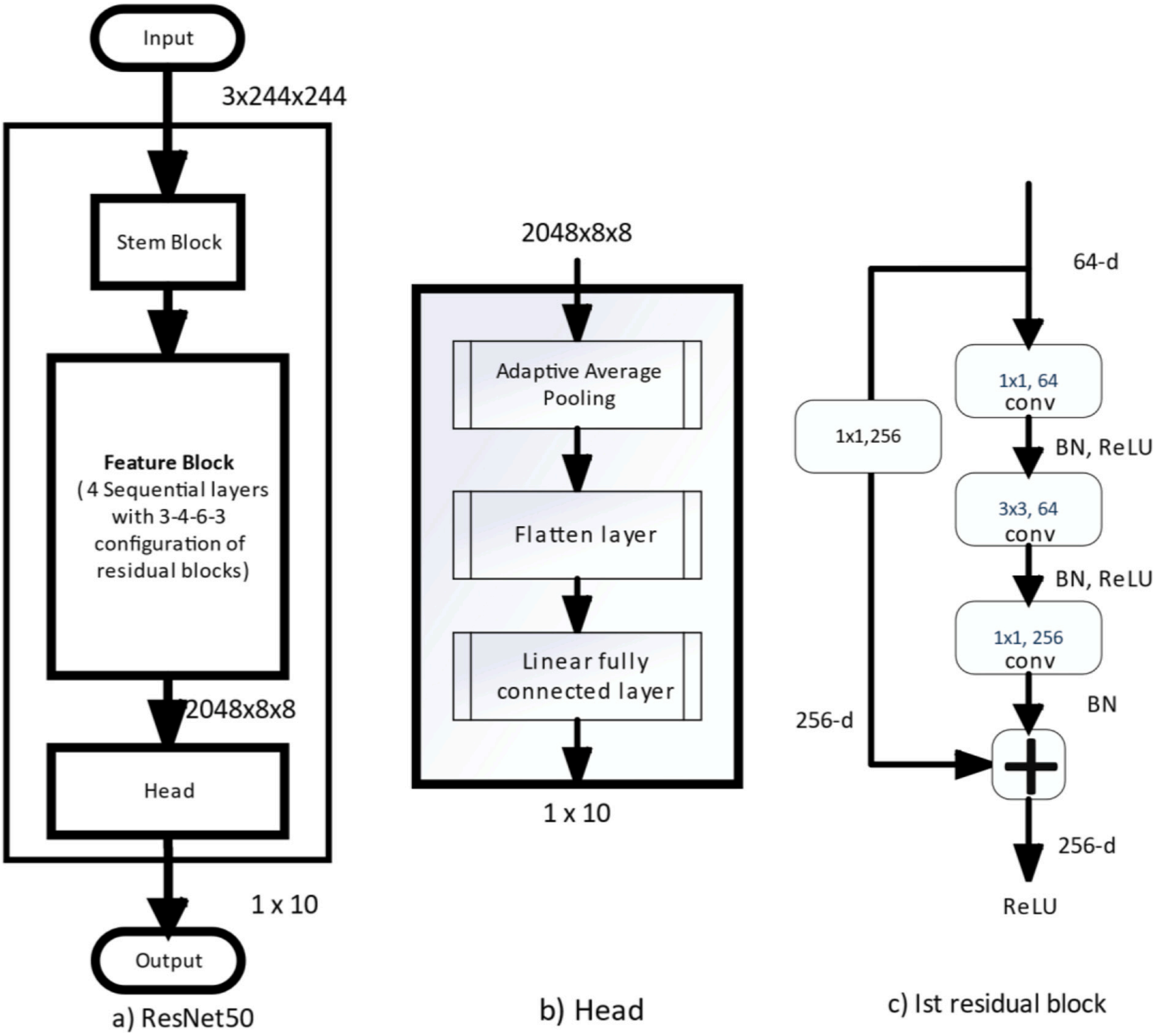


Fig. 5. ResNet Network block diagram. a) is the overall block architecture. b) shows zoomed in view of the architecture of a ResNet head. c) shows the 1st residual block with 3 stages of convolution (each followed by batch normalization, and ReLU) and the residual path. The filter channel changes hierarchically from 64, 256, 512, 1024, 2048 in the last sequential layer.

based deep neural network architecture that can be used for image classification and as an encoder for image segmentation tasks. The architecture mainly consists of a stem or patchify block, a feature block, and a fully connected linear head. The feature block consists of four sequential layers with 2, 2,6, and 2 Swin-Transformer-V2 blocks in the first, second, third, and fourth layer, respectively (see Fig. 4). The vision transformers differ from convolutional neural networks in that they divide the input image aggressively into non-overlapping regions in the stem block and pass the patches after flattening it into one dimension. The patchification is performed using a patch embedding in the stem block, which is a convolutional layer, followed by layer normalization and flattening in the stem block. Each of the four sequential layers is connected by a patch-merging layer. The swin2-transformer block is more complex than the ConvNeXt block or the residual bottleneck block, as shown in Fig. 4. The swin2-transformer block first has a weighted multi-head self-attention layer with a kernel size of 7 × 7 with a skip connection. It is followed by an MLP layer and a layer-normalization layer with a skip connection. The filter channel varies from 128, 256, 512, and 1024 in the four sequential layers. The output for architecture is a 144 × 1024 feature map (see Fig. 7). A linear, fully connected layer is connected to develop an image classifier, ‘S’ models.

2.3. *Model training and empirical evaluation*

We used Python on Linux operating systems, on a system with a GTX-3060 NVIDIA GPU for Training and evaluating the models. For segmentation, we used the MMSegmentation [80] an opensource libraries by OpenMMLab [81] based on PyTorch [80], for the classification PyTorch-**Im**age-**M**odels(timm) [82] libraries were used. Both libraries provide a collection of state-of-the-art architectures, pre-trained models for deep-feature extraction, and basic building blocks of deep learning architecture, such as layers, utility functions, optimizers, schedulers, data loaders, and augmentation. We used such libraries to provide a reference implementation for evaluation and reproduction. We used Weights & Biases [83] for pavement rating classification experiments. It provides experiment tracking, dataset and model versioning, hyper-parameters, and data visualization. The backbone of a neural network (N.N.), also called encoder for segmentation or feature block in classification N.N., requires extensive pre-training on image benchmarks to learn the filters’ weight and biases (also referred to as model parameters) for extracting distinctive features from images. ImageNet [84] has been widely used for pre-training deep learning architectures. It contains 1000 image classes, while ImageNet-21 K [85] contains 21,000 image classes. These pre-trained parameters are provided by many open-

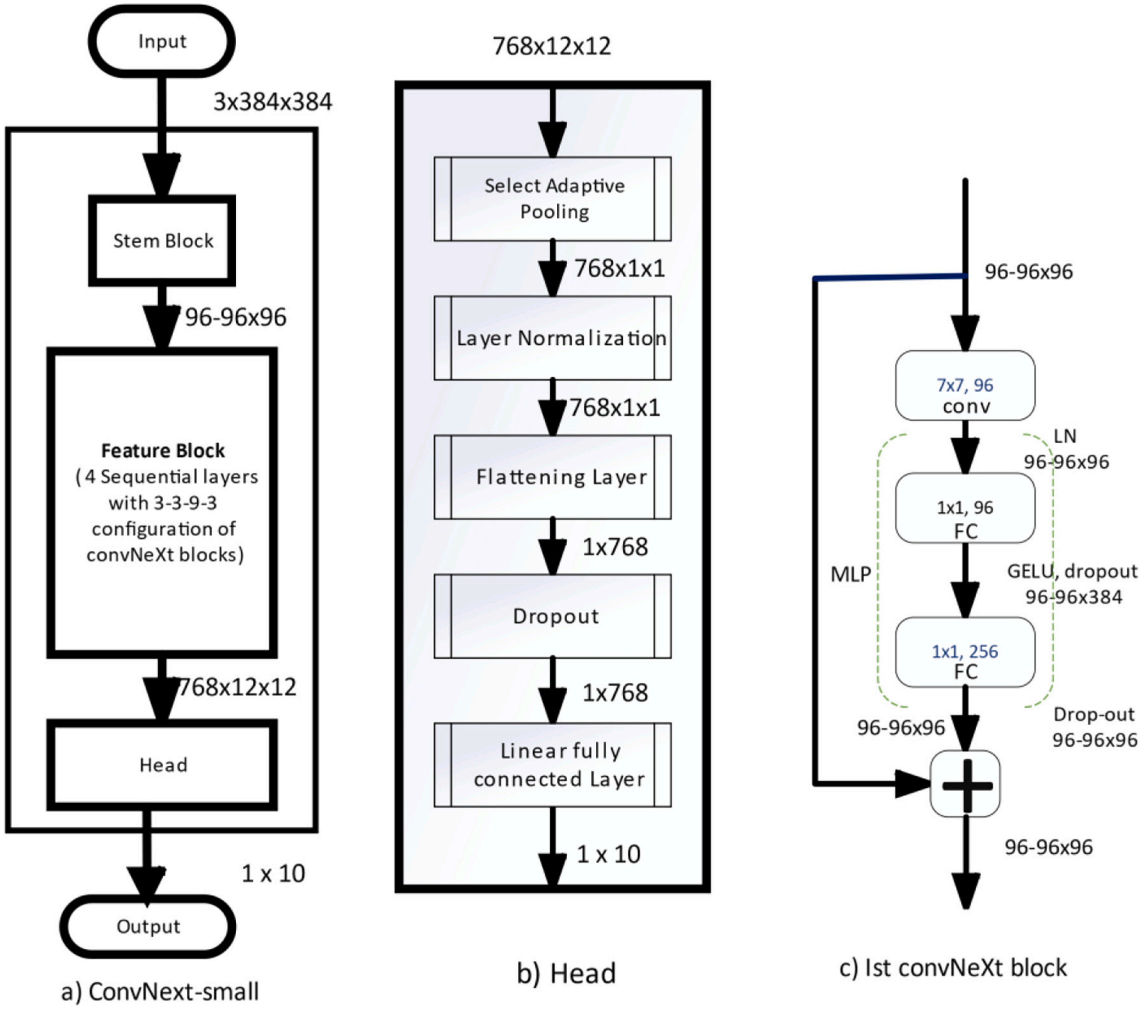


Fig. 6. ConvNeXt Network block architecture. a) is the overall architectural design. b) show zoomed in view of the architecture of the head. c) Show the 1st ConvNeXt-small block in the 1st sequential layer. The filter channel changes hierarchically in the sequential layers.

source libraries such as MMClassification [86], PyTorch-Image-Models [87], PyTorch-official model zoo [88], TensorFlow Model Zoo [89], and Model-Zoo [90]. We first explain the training parameters for image segmentation and then image classification tasks.

2.3.1. *Image segmentation*

A total of three models were developed and then evaluated to choose the best semantic (object) segmentation model for our task. The choice of the deep-learning architectures used is based on benchmark dataset performance tabulated by [91], memory required by the model to load, and the inference time (i.e., frame per second). Table 4 summarizes the three deep learning architectures used to develop the models, the pre-trained backbone used, memory requirement, and inference time.

First, two models were developed by training a modified DeeplabV3 architecture (with a separate auxiliary head for better optimization) for seven segmentation classes; one with ResNet50 [71] and the other with ResNet50b [71,72] encoder (backbone). The stem block has a stride of 8, i.e., eight times down sampling from 512 × 512 to 64x64x64. The padding configuration used is 1,1,2,4, and strides are 1,2,1,1 for the 1st, 2nd, 3rd, and 4th sequential layers, respectively. The output of the encoder is a 1 × 2048 channels feature map. During training, the fully connected auxiliary head block is optimized using a SoftMax loss function. We used the cross-entropy loss function for the decoder with a dropout ratio of 0.1 in the auxiliary head and the main decoder head, ASPP. A pre-trained model [82] initializes weight and biases for encoders ResNet50 and ResNet50b. The decoder and the auxiliary heads were initialized with random normal distribution values with a standard deviation = 0.01.

Second, one model was developed by training standard segFormer architecture. The encoder has four sequential layers with an overlapped patch embedding of 7, 3, 3, and 3, with the layers number of each transformer encode-layer as 3, 4, 6, and 3. The stride of each overlapped patch embedding is 4, 2, 2, and 2, with a spatial reduction rate of each transformer, encode layer to be 8, 4, 2, and 1—the encoder's drop path rate = 0.1 and the dropout rate 0.1. The loss function used in the decoder is cross-entropy loss. A pre-trained model [82] is used to initialize weight and biases for the encoder. The decoder head is initialized with random values of a normal distribution with a standard deviation = 0.01.

The input size of the image dataset is 720 × 576 × 3, while the input size of the model is 512 × 512 × 3. Therefore, an online training augmentation pipeline is used containing resize (size = 512 × 512, ratio 0.5 to 2.0), random-crop (max-ratio = 0.75), random-flip (ratio = 0.5), photometric distortion, and normalization (with ImageNet standard deviation and mean) layers. We used stochastic gradient descent (SGD) for DeepLabV3 and Adam weight-decay for SegFormer as an optimization function with an initial learning rate = 0.01, learning rate decay policy = 'polynomial' with power = 0.9, momentum = 0.9, and weight_decay = 0.0005. The training was conducted for 160,000 iterations for all three models trained.

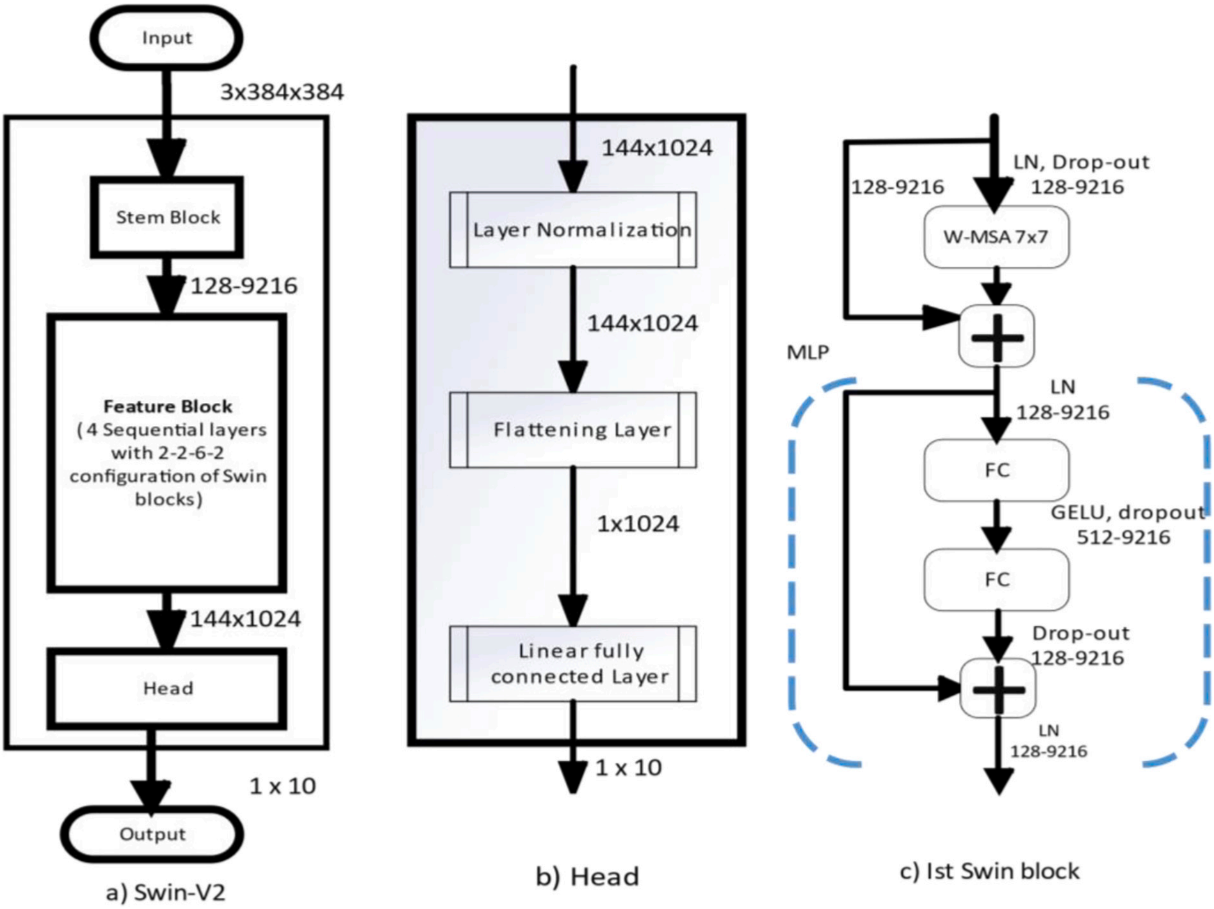


Fig. 7. SwinV2-transformer Network block diagram. a) The overall block summary. b) show zoomed in view of the architecture of. Head. c) The 1st Swin-V2 building block of the sequential feature block

Table 4
Model used for evaluating the semantic segmentation algorithm for pavement pixel extraction.

Method	Pretrained-backbone	Mem (GB)	Inference time (fps)
DeepLabV3	R-50-D8	6	2.74
DeepLabV3	R-50b-D8	6	2.74
Segformer	MIT-B2	7.42	3.36

2.3.2. PSCI rating classification

As mentioned in the methodology, three architectures were developed and evaluated for pavement rating classification—one with the baseline ResNet50 (R1, and R2 models), the second with ConvNeXt (C1, C2, and CT1 models), and the third with SwinV2-base (S1, and S2 models) architectures. A list of seven developed models is shown in Table 5, which summarizes the type of input images used, size of images, total parameter counts in millions, batch size, number of the epoch, learning rate, momentum, and optimizer. The stem and feature block were initialized with the pre-trained mode [87]. The weights and biases of the classification head and the rest of the trainable model parameters are then retrained using the RoadSurvey-2 dataset with their ground truth labels. The images are to be classified for different pavement surfaces condition, which means the model should be trained to make a classification decision based only on the pavement pixels i.e. images

Table 5
List of seven developed models including the summary of the type of input images used, size of images, total parameter counts in millions, batch size, number of the epoch, learning rate, momentum, and optimizer.

Models	version-name	Input image type	Size of image	P-count (Millions)	Batch size	Epoch	lr	momentum	optimizer
resnet50									
resnet50	R1	Segmented	224	25.56	32	300	0.0004	0.9	SGD
resnet50	R2	Original + Segmented							
swinv2_base									
swinv2 base	S1	Segmented	384	87.92	4				SGD
swinv2 base	S2	Original + Segmented			2				
convNeXt									
convNeXt-small	C1	Segmented	384	50.22	16				adamw
convNeXt-small	C2	Original + Segmented							
convNeXt-tiny	CT1	Segmented		28.59					

with pavement pixels extracted. On the other hand using the original image (the image with pavement as well as the rest of the pixels) along with its corresponding pavement extracted images for training the classification model should increase the robustness to varying background pixels. To test this hypothesis, inspired by background augmentation technique by [92], we developed models with different input image types. In column three of Table 5 i.e., 'Input Image' the word 'Segmented' means only images with pavement pixels extracted, and cropped are used for training, while 'Segmented + Original' images means, the segmented images and its corresponding cropped original images is used for training the model. This means models (R2, S2, C2) that use segmented plus original images have twice the number of training images than models that use only segmented images (R1, S1, C1, and CT1). The images were resized to either 224 for R models and 384 for C and S models before passing on to the models both for training and inference.

2.3.3. Evaluation metrics

The segmentation model's evaluation criteria are the pixel accuracy and the intersection over union (IoU) for each of the seven classes and the mean value of accuracy and IoU.

The per-class pixel accuracy (see Eq 2) is evaluating a binary mask, i.e., a true positive pixel represents a pixel correctly predicted to the given ground truth class. Whereas, a true negative pixel represents a pixel that is correctly predicted as not from the given class.

$$Accuracy = \frac{True\ positive + True\ negative}{True\ positive + False\ positive + True\ negative + False\ negative} \tag{2}$$

The mean pixel accuracy is the average pixel accuracy of each class. The IoU metric for each class quantifies the overlap between the predicted mask and the ground truth mask and is calculated as given in the equation Eq 2

$$IoU = \frac{predicted\ mask \cap groundtruth\ mask}{predicted\ mask \cup groundtruth\ mask} \tag{3}$$

The evaluation criteria for classification are precision, recall, and F1-score per class, and the F1-score across all classes is computed using the equations Eq 4, 5, 6 and 7 below.

$$precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$Average\ Recall = \frac{\sum_{i=10} TP_i}{Total\ No.of\ Images} \tag{6}$$

$$F1\_score = 2{*}\frac{precision{*}recall}{precision + recall} \tag{7}$$

where TP is a true positive, FP is a false positive, TN is a true negative, and FN is a false negative.

We also include the weighted Cohen's Kappa score [93,94] for each model developed compared to the ground-truth labels assigned by expert data analysts for PSCI ratings. A confusion matrix is created for PSCI ratings between two raters, in which an element f_{ij} represent a number of images classified as a category i by rater-1 and category j by rater-2. The elements where $i = j$, are the agreements between two PSCI data analyst and $i \neq j$ is the disagreement between the two raters. While r_i and c_j the row and column totals for category i and j. Then, the weighted Cohen's Kappa gives a measure of agreement and degree of disagreement between two raters and is given by Eq. Eq 8 [95].

$$\kappa_w = \frac{P_{o(w)} - P_{e(w)}}{1 - P_{e(w)}} \tag{8}$$

where, $P_{o(w)} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij} f_{ij}, P_{e(w)} = \frac{1}{N^2}\sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij} r_i c_j,\ \ w_{ij} = 1 - \frac{|i-j|}{k-1.}$

3. Results and discussion

This section reports the quantitative results for semantic segmentation and PSCI classification on the respective test sets. We also discuss some strengths and limitations of the automated PSCI rating approach.

3.1. Semantic segmentation

Two different architectures were evaluated for image segmentation block. Table 6 summarizes the quantitative results of the three semantic segmentation models developed and evaluated for seven classes. The results show that the model developed using DeepLab version-3 with ResNet50-B performed best, with a mean IOU of 75.71% and a pixel accuracy of 82.05%. The segFormer performance was poorer than that of the DeepLab model. The pixel accuracy for classes road, human, and vehicles for DeepLab-RestNet50-b, which are important for the automated PSCI classification application, are 97.74%, 85.25%, and 94.98%, with an IOU of 94.76%,76.18%, and 91.49% respectively. The IOU for poles, traffic lights, and traffic signs is poorer than the rest of the classes because of the coarse labelling in the IrishRoadSurvey-1. The other reason is that the poles for light-pole, traffic-light, and traffic-sign are sometimes visually similar due to colour, shape, and texture. In Table 6, we observe that segFormer [96] did not perform better than DeepLab on our evaluation dataset with seven classes which is due to the low pixel accuracy and IOU for the four classes, i.e. human, pole, traffic light, and traffic sign, for the SegFormer [96] based model. Fig. 10: show sample, from test dataset, of segmentation mask added on top of original images generated using the model DeepLab-RestNet50-b. We conclude that the low performance is because of the low-resolution image we used to train the model, which can also be concluded from similar experiments presented by [97], and due to the coarse ground-truth labelling for the referenced classes in the IrishRoadSurvey-1 dataset.

3.2. PSCI rating classification

For a fair comparison we compared all the seven models with the ground truth. The ground truth is the data labelled by the expert data analyst-1. Fig. 8 show the graph between the validation accuracy, and training loss w.r.t. the epoch during the training of all the seven models on three different architectures. We include the F1-Score graph of the second data expert (human) along with other seven models as shown in the Fig. 9. The weighted Cohen kappa score between two expert data analysist is 0.98, while the average F1-score was 0.97 with an average precision and recall of 97% (a more detailed comparison is shown above in Table 3).

Fig. 9 shows the F1-scores of different automated models developed in this study for PSCI classification on the test dataset of Roadsurvey-2. The results on the test data show that the automated system in general has a very good F1-score for ratings 1,6, and 10, while low F1-scores can be observed for classes 3 and 4 in general.

A summary of evaluation results of seven different models is presented in Table 7. The models R2, S2, and C2 use segmented plus original images, while the models that use only segmented images R1, S1, C1, and CT1 use segmented images.

For each model, we show class wise F1-score for test images with and without pavement segmentation. The column labelled 'Seg' mean result of the segmented images. The column labelled 'Orig' means test result without applying segmentation. The comparison between The mean F1-score, the mean precision, recall, and Cohen Kappa score of each model is presented. The values in bold are the highest. We see that S2 (developed using SwinV2 using both segmented and original images) and C2 (developed using ConvNeXT using both segmented and original images) have better performance than the rest of the models. Models that are

Table 6
Summary of evaluation of the different segmentation algorithms used for evaluating pavement segmentation. The evaluation is based on pixel accuracy and Intersection over union (IOU).

Class	Iterations	160,000		160,000		160,000	
		Test dataset pixel accuracy and IoU					
	Classes	segFormer (Transformer+MLP decoder)		Deeplab ResNet-50_v1		Deeplab ResNet-50b	
		IOU	Accuracy	IOU	Accuracy	IOU	Accuracy
0	BACKGROUND	90.67	95.91	93.88	97.89	94.76	97.73
1	Human	50.11	58.74	73.59	82.8	76.18	85.25
2	Pole	18.87	20.49	45.91	52.11	50.3	58.41
3	Road	92.92	96.94	94.73	96.44	95.71	97.74
4	Traffic Light	11.08	11.73	54.03	63.97	60.22	70.19
5	Traffic Sign	24.17	27.76	57.82	69.69	61.31	70.01
6	Vehicle	70.74	88.3	90.63	94.92	91.49	94.98
	Avg/mean	52.65	57.12	72.94	79.69	75.71	82.05

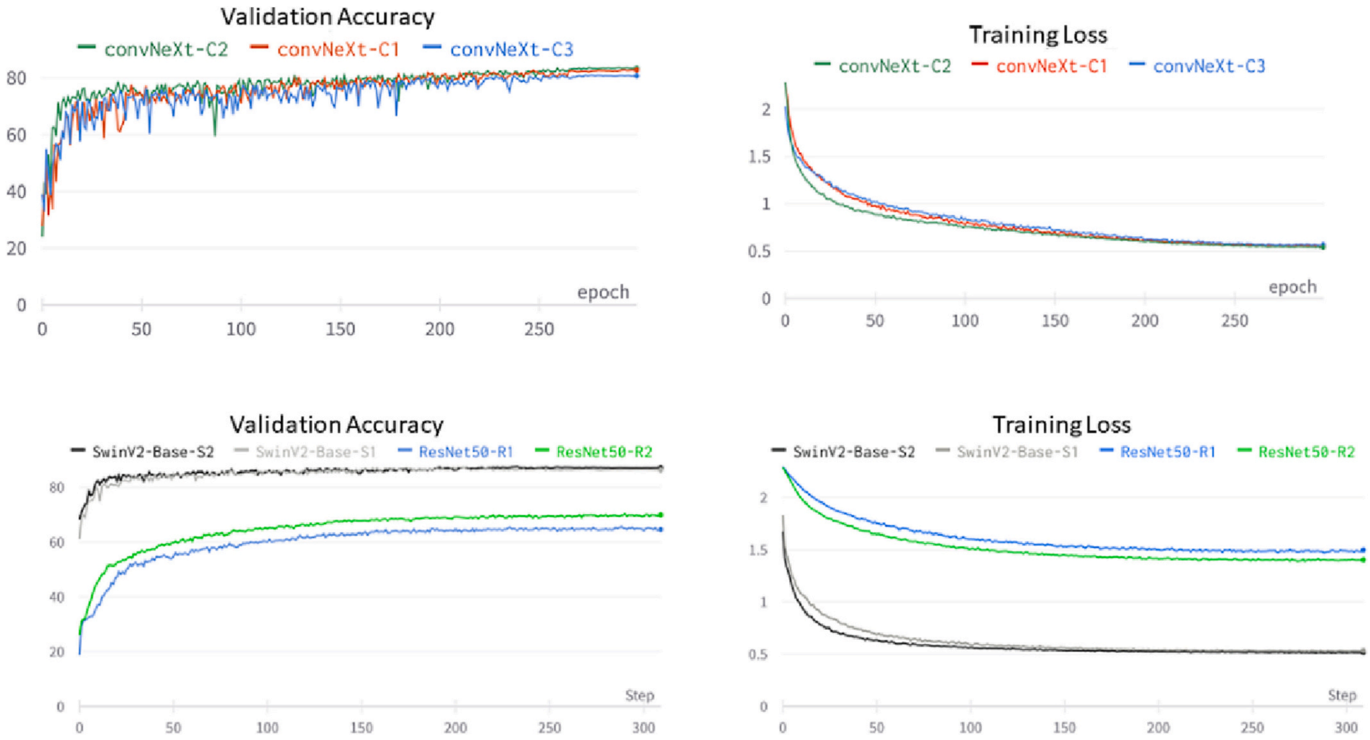


Fig. 8. Graph between the validation accuracy, and training loss w.r.t. the number of epoch during the training of all the seven models.

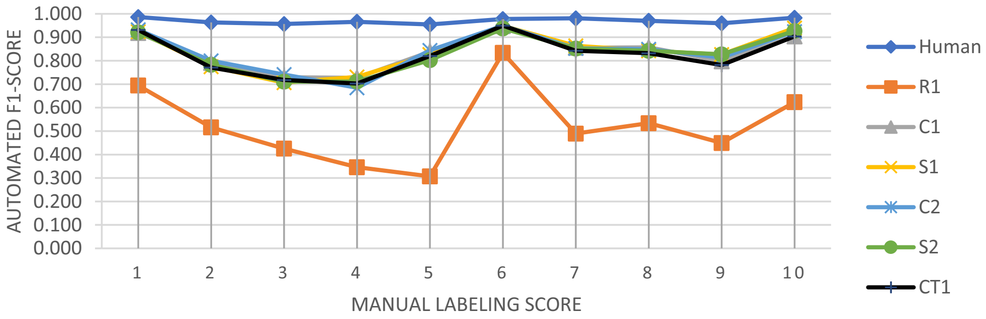


Fig. 9. Comparison between the F-Score for different PSCI ratings method including the a typical human analyst for the test data in RoadSurvey-2 dataset.

trained on the original and the segmented images have performed slightly better than models that use only segmented images. Moreover, this also show that both the models' S′ and 'C' models in general are learning the PSCI ratings from pavement pixels rather than background pixels. This also show that the models trained using this set of images are more robust to background when compared to models that are trained only on pavement segmented images. It can be concluded that 'S' and 'C' models are better performing because of the patchification performed

Table 7
Summary of seven different models developed using three different archiectures and 2 different image augmentation techniques. It summarizes the class wise F1-score, the mean F1-score, precision, recall, and weighted Kappa sore. The values in the bold are the highest.

	F1-Score													
Model	R1		R2		S1		S2		C1		C2		CT1	
PSCI	Seg	Orig	Seg	Orig	Seg	Orig	Seg	Orig	Seg	Orig	Seg	Orig	Seg	Orig p
1	0.694	0.229	0.750	0.776	0.924	0.709	0.921	**0.949**	0.919	0.805	0.932	0.918	0.931	0.780
2	0.516	0.474	0.580	0.583	0.775	0.605	0.783	0.791	0.791	0.633	0.800	**0.812**	0.771	0.585
3	0.426	0.456	0.445	0.449	0.707	0.582	0.711	0.738	0.730	0.599	0.741	**0.761**	0.718	0.581
4	0.346	0.248	0.400	0.387	**0.731**	0.647	0.712	0.726	0.727	0.628	0.685	0.689	0.703	0.595
5	0.307	0.056	0.325	0.264	0.826	0.697	0.803	0.831	0.838	0.711	0.844	**0.849**	0.820	0.659
6	0.833	0.652	0.846	0.716	0.949	0.923	0.938	0.944	0.948	0.925	0.949	**0.954**	0.950	0.938
7	0.489	0.353	0.559	0.541	**0.864**	0.784	0.852	0.857	0.855	0.746	0.853	0.859	0.843	0.745
8	0.534	0.395	0.603	0.634	0.843	0.818	0.842	0.845	**0.857**	0.793	0.849	0.847	0.832	0.787
9	0.449	0.340	0.492	0.496	**0.827**	0.765	**0.827**	0.815	0.798	0.717	0.813	0.804	0.782	0.684
10	0.624	0.587	0.667	0.712	**0.940**	0.905	0.927	0.925	0.904	0.897	0.925	0.930	0.904	0.866
Avg-F1	0.522	0.379	0.567	0.556	0.838	0.744	**0.831**	**0.842**	0.837	0.745	**0.839**	**0.842**	0.825	0.722
Avg-Precision	0.543	0.495	0.585	0.571	0.837	0.770	0.828	0.841	0.839	0.777	0.839	0.842	0.828	0.762
Avg-Recall	0.531	0.393	0.571	0.566	0.859	0.781	0.852	0.860	0.858	0.782	0.859	0.862	0.848	0.765
weighted-Kappa	0.617	0.314	0.668	0.689	0.887	0.816	**0.884**	**0.886**	0.882	0.804	**0.883**	**0.886**	0.876	0.779

before the sequential feature block as compared to 'R' models that uses aggressive downscaling of the images.

Table 8 shows the comparison between recall and precision for the 10 PSCI ratings between the C2 and the S2 model. The Table 8 show that the S2 model has a higher precision for a given recall than C2. Model-C2 can predict a PSCI rating with an average recall of 84.10% ±8.60% and with a precision of 83.90% ± 8.17% between 1 and 10 PSCI classes. Model-S2 can predict a PSCI rating with an average recall of 83.60% ±8.30% with a precision of 82.85 ± 8.10. Fig. 11:shows sample false postive by Model C2 for the PSCI ratings 1 and 10.

We see that S2 model built with a transformer backbone and C2 model with a convolutional neural network backbone are comparable. The model C2 has 42% less trainable parameters than S2 and is therefore much faster to retrain on new data than S2. Table 9 shows the confusion matrix on the test dataset for the models S2 and C2. The cells along a row show the predicted class for a manual PSCI rating classification. The empty cell means the value is zero. The diagonal cells show the true positive rate, while the cells along a row of a manual PSCI rating show the false negative rate; the cells along a column of a predicted PSCI rating except the diagonal element show the false positive rate. We observe some randomness in prediction specially for the middle classes i.e., 3, 4, and 5. We see that the false prediction is mostly of the adjacent classes, which is a natural disagreement occur even by the human data analyst. The type of distress present in these middle classes varies from linear cracking, patching, alligator cracking, and potholes, which may vary in shape, size, and texture. It is also important to analyse if the position of these distresses with respect to the camera effect the predicted ratings by the model. The comparison of both S2 and C2 show that the predicted output is noisier for S2 than the C2 model. Table 10 show the confusion matrix between manual PSCI rating to predicted PSCI rating, recall, and F1-score when the adjacent classes are combined for a 5-class classification. The choice of adjacent classes is based to relate the five treatment measures as given in the PSCI rating manual and shown in Table 1. The recall for each class is >90% except of the class three-four; similarly, the F1-score, which is the harmonic mean of precision and recall is higher than 0.9 for all except for class three-four. This indicates that the developed models hold promise towards an automated PSCI system that is related to treatments measure.

Table 8
Recall and precision of Models C2 and S2 for 10 PSCI ratings.

	Recall		Precision	
PSCI	C2	S2	S2	C2
1	**0.948905**	0.941606	0.902098	**0.915493**
2	**0.814815**	0.8	0.765957	**0.785714**
3	**0.767857**	0.702381	**0.719512**	0.716667
4	0.672619	**0.690476**	**0.734177**	0.697531
5	0.797386	**0.836601**	0.771084	**0.897059**
6	**0.934164**	0.907473	**0.969582**	0.965074
7	**0.868852**	0.860656	**0.843373**	0.837945
8	0.871886	**0.893238**	0.796825	**0.827703**
9	0.781818	**0.804545**	**0.850962**	0.847291
10	**0.951456**	0.92233	**0.931373**	0.899083

4. Conclusions

In conclusion, our study successfully developed a deep learning framework for automated pavement rating based on the PSCI standard for regional and local roads. However, there are some limitations to our study. Firstly, we only used data from the Irish road network, so it remains to be seen whether our model can be generalized to other regions. Secondly, while we developed a benchmark dataset for PSCI rating and a labelled dataset for pavement segmentation, it is still possible that our training dataset does not capture the full range of distresses associated with different classes of PSCI rating across road networks. While we made efforts to include as many variations as possible, we acknowledge that there may still be some diversity of distresses that we did not capture. The dataset was subjected to a cleaning process to eliminate images that exhibited motion blur, insufficient lighting, and focus blur. This is a standard practice in many domains that use machine learning, as low-quality data can lead to suboptimal models. Notably, the framework incorporates an image processing step, which can effectively filter out such images from any similar dataset. In our framework, we utilize a direct classification approach for predicting PSCI ratings. While this approach offers high accuracy, it can be considered a "black box" method that may limit interpretability. However, from discussions with domain experts, accuracy is more important than interpretability in the pavement rating task, given the low-risk nature of the task. The interpretability of the model can be improved by combing its use with the PSCI rating manual (see Table 1) which gives the summary of each PSCI rating with regard to distress indicators, condition of surface and structure, and the treatment measures required for each rating. Therefore, given a PSCI rating manual and the direct PSCI rating value, the authorities can relate a treatment measure the patch require.

Regional and local pavements cover a major part of the country's road network, especially in Ireland. The most economical data collection for regional and local roads is from a video van (including an onboard computer to record distance and position) with a camera mounted on the front of the dashboard giving the frontal view. Our deep learning framework for PSCI classifications is divided into three main components, 1) a pavement pixel segmentation block, 2) an image processing

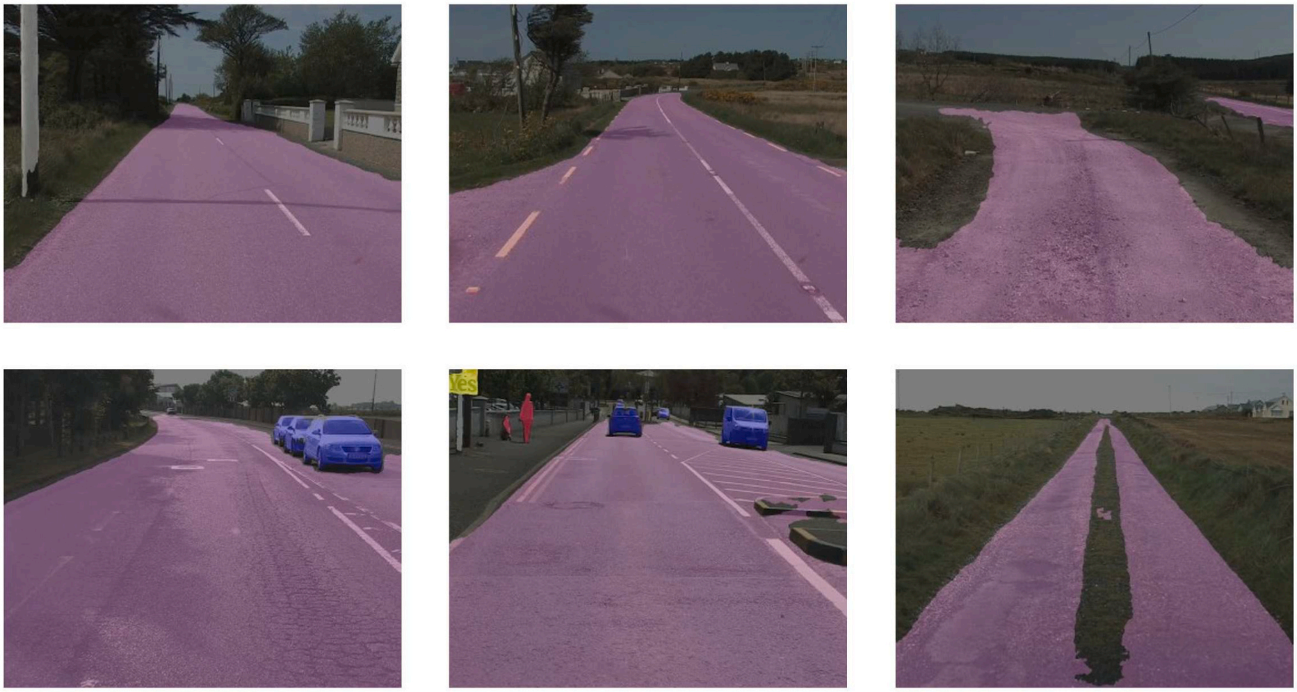


Fig. 10. Sample segmentation mask overlaid on original images from test dataset using the Model DeepLabV3 b trained on IrishRoadSurvey-1 and Cityscape dataset with 7 classes.

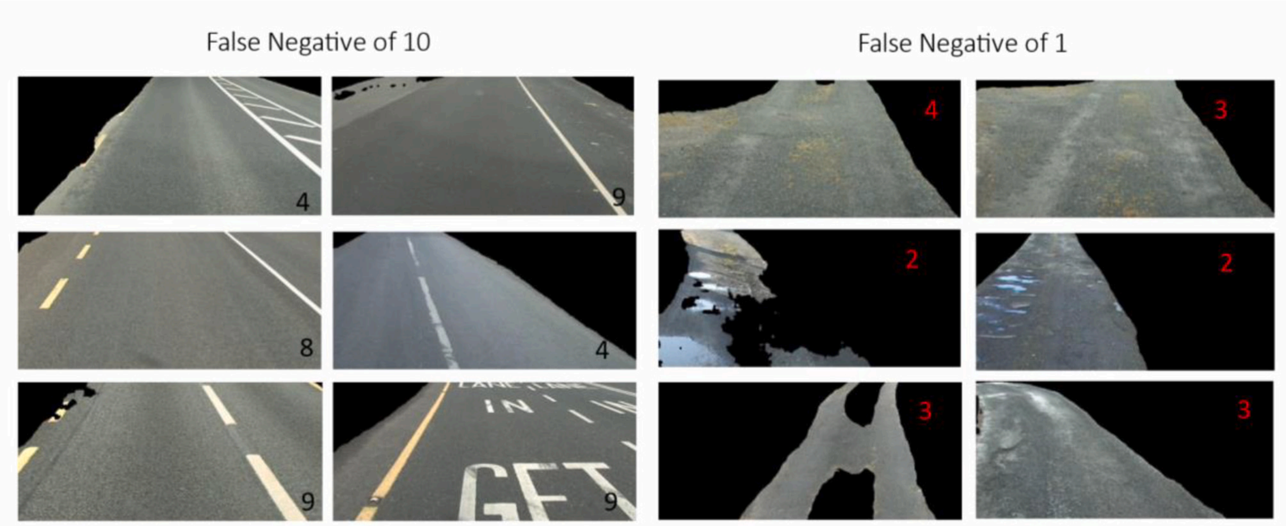


Fig. 11. Sample False negative for class 10 and for class 1 using the Model C2. The number inside the image gives the predicted label of the images.

block, and 3) a PSCI-rating classification block. We developed a benchmark dataset for PSCI rating containing 7453 images labelled for 1–10 PSCI. We also developed a labelled dataset for pavement segmentation containing 500 images to capture statistics on regional and local pavements in Ireland. We developed three models for the first blocks and evaluated their performance against each other using our dataset RoadSurveyDataset-1. For the second component, we developed an algorithm to remove poor segmented images, which improved the PSCI classification accuracy when compared to previous results [60]. For the third component, we developed seven models using three different architectures (ResNet50, ConvNeXt, SwinV2) and evaluated their performance against each other using our PSCI benchmark dataset.

The models trained with our unique augmentation technique, i.e., training on the cropped and segmented pavement image and its corresponding cropped original image, have resulted in a better performance than the models that only use segmented pavement images. Overall, the models using ConvNeXt, and SwinV2 performed better than the baseline ResNet50 architecture-based models. Model-C2 can predict a PSCI rating with an average recall of 84.10% ±8.60% with a precision of 83.90% ± 8.17% between 1 and 10 PSCI classes. Model-S2 can predict a PSCI rating with an average recall of 83.60% ±8.30% with a precision of 82.85 ± 8.10. It can be concluded that 'S' and 'C' models perform better because of the patchification performed before the sequential feature block than 'R' models that use aggressive downscaling of the images.

It is important to note that the models were trained and evaluated on data collected from a video van with a camera mounted on the front

Table 9

Confusion matrix for PSCI classes for classifiers trained on SwinV2 (S2) and ConvNeXt (C2) models using segmented and original images. Each column shows the percentage of test images (segmented plus original) classified in that class. The empty cell means zero. The PSCI rating class in the row is manual rating, while the predicted PSCI class is in the column.

Confusion Matrix for 10 PSCI ratings using C2 Model

PSCI	1	2	3	4	5	6	7	8	9	10
1	94.5%	2.6%	2.9%							
2	5.6%	80.7%	12.6%				1.1%			
3	1.5%	10.1%	77.7%	8.6%	0.6%		0.9%		0.6%	
4	1.2%	1.8%	12.2%	67.9%	1.5%	3.9%	5.4%	2.7%	3.6%	
5	0.3%	1.3%	2.9%	4.2%	80.4%	3.6%	4.2%	1.3%	1.0%	0.7%
6		0.1%	0.1%	1.3%	0.5%	93.6%	0.7%	2.5%	0.8%	0.4%
7	0.4%	0.2%	0.6%	3.3%	0.8%	0.4%	87.1%	5.7%	1.4%	
8				3.0%	0.7%	0.5%	3.6%	87.5%	3.6%	1.1%
9				1.6%	1.6%	0.9%	3.0%	8.4%	77.3%	7.3%
10			0.5%	0.7%	0.2%	0.5%	0.5%		1.9%	95.6%

Confusion Matrix for 10 PSCI ratings using S2 Model

PSCI	1	2	3	4	5	6	7	8	9	10
1	94.53%	4.01%		1.09%	0.36%					
2	4.07%	80.74%	12.96%	0.37%	0.37%		1.11%		0.37%	
3	1.49%	12.80%	71.13%	10.42%	0.89%	0.60%	1.79%	0.30%	0.30%	0.30%
4	0.30%	0.89%	11.31%	70.83%	3.27%	1.79%	5.65%	5.65%	0.30%	
5	0.33%	1.63%	1.96%	3.92%	83.66%	2.29%	3.92%	0.33%	0.65%	1.31%
6				1.16%	2.67%	91.46%	0.62%	2.49%	1.25%	0.36%
7	0.61%	0.82%	0.61%	1.84%	1.64%	0.41%	86.07%	6.97%	1.02%	
8			0.36%	1.42%	0.36%	0.71%	3.38%	89.68%	3.38%	0.71%
9			0.23%	1.14%	2.05%	1.82%	2.05%	9.32%	79.32%	4.09%
10				0.49%		1.21%		1.21%	4.37%	92.72%

Table 10

Recall and precision, and confusion matrix for C2 and S2 model if we combine the adjacent classes as shown in the table.

Confusion Matrix after merging adjacent classes using C2 Model

5-Class	One-Two	Three-Four	Five-Six	Seven-Eight	Nine-Ten	Recall	F1-Score
One-Two	91.7%	7.7%		0.6%		91.7%	0.91
Three-Four	7.3%	83.2%	3.0%	4.5%	2.1%	83.2%	0.82
Five-Six	0.4%	2.7%	92.0%	3.7%	1.3%	92.0%	0.94
Seven-Eight	0.3%	3.4%	1.2%	91.9%	3.1%	91.9%	0.90
Nine-Ten		1.4%	1.6%	6.1%	90.8%	90.8%	0.92

Confusion Matrix after merging adjacent classes using S2 Model

5-Class	One-Two	Three-Four	Five-Six	Seven-Eight	Nine-Ten	Recall	F1-Score
One-Two	91.7%	7.2%	0.4%	0.6%	0.2%	91.7%	0.90
Three-Four	7.7%	81.8%	3.3%	6.7%	0.4%	81.8%	0.83
Five-Six	0.4%	2.2%	92.4%	3.4%	1.7%	92.4%	0.94
Seven-Eight	0.7%	2.1%	1.5%	93.0%	2.7%	93.0%	0.90
Nine-Ten		0.9%	2.6%	6.5%	90.0%	90.0%	0.92

dashboard, which may not accurately represent the conditions encountered by engineers in the field due to imaging sensor quality, depth of view, camera angle, height from the surface, and weather condition. Further evaluation using real-world video frames of stretches of regional and local roads will be necessary to assess the practicality and usability of our approach. Additionally, our study highlights the need for more research in this field, as there is very little academic literature available for direct pavement rating estimation. Our approach can be more easily applied in practice as it does not need dataset with individually labelled distresses, rather only dataset with ratings applied to each image which are easier to obtain as it is practiced in real-world in this domain. Our work can serve as a starting point for researchers to develop automated rating frameworks from image and video data for other regions that rate pavements using a standard rating scale based on visual distresses.

Funding

Dr. Waqar S. Qureshi has received funding under the Marrie Currie Career-Fit plus postdoctoral fellowship program from Enterprise Ireland Grant No. M.F. 2021 0273, partially funded by the European Union's Horizon 2020 Research and Innovation Program under the Marie Sokolowski-Curie Co-funding of regional, national, and international programs Grant agreement No: 847402.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Dr. Waqar S. Qureshi has received funding under the Marrie Currie Career-Fit plus postdoctoral fellowship program from Enterprise Ireland Grant No. M.F. 2021 0273, partially funded by the European Union's Horizon 2020 Research and Innovation Program under the Marie Sokolowski-Curie Co-funding of regional, national, and international programs Grant agreement No: 847402.

David Power, Brian Mulry, and Kieran Feighan work at Pavement Management Services Private Limited, which has provided PSCI labelled images for pavement surfaces in Ireland. They have contributed to the conception, acquisition of data, validation, and interpretation of data.

Waqar S. Qureshi and Dympna O'Sullivan wrote the original draft. Waqar S. Qureshi, Dympna O′ Sullivan, and Ihsan Ullah have contributed to the conception and design of experiments, data analysis, and interpretation of data. Ihsan Ullah and Susan McKeever have been involved in drafting the manuscript, proofreading, and revising it critically for important intellectual content. The authors have no potential competing interests.

Data availability

The datasets and the results that supported the findings of this study and were analyzed during the current study are available from the corresponding author at reasonable request.

References

[1] Road Pavement Surface Types, 2022, p. 1. https://interpro.wisc.edu/tic/?csis-search-options=site-search&s=paser&submit=Search (accessed February 3, 2022).

[2] I. Government, Local Authority Budgets 2021. https://assets.gov.ie/139273/8554c7e7-d87c-4185-8cc1-32c8bf51c5c3.pdf, 2021.

[3] W.S. Qureshi, S.I. Hassan, S. McKeever, D. Power, B. Mulry, K. Feighan, D. O'Sullivan, An exploration of recent intelligent image analysis techniques for visual pavement surface condition assessment, Sensors 22 (2022) 9019, https://doi.org/10.3390/S22229019.

[4] John S. Miller, William Y. Bellinger, Distress Identification Manual for the Long-Term Pavement Performance Program, Georgetown Pike. https://highways.dot.gov/sites/fhwa.dot.gov/files/docs/research/long-term-pavement-performance/products/1401/distress-identification-manual-13092.pdf, 2014 (accessed March 24, 2022).

[5] PASER Asphalt Roads Pavement Surface Evaluation and Rating PASER Manual Asphalt Roads. http://tic.engr.wisc.edu, 2002 (accessed February 3, 2022).

[6] N.S.P. Peraka, K.P. Biligiri, Pavement asset management systems and technologies: a review, Autom. Constr. 119 (2020), 103336, https://doi.org/10.1016/j.autcon.2020.103336.

[7] Standard Practice for Roads and Parking Lots Pavement Condition Index Surveys. https://www.astm.org/d6433-09.html, September, 2011 (accessed April 16, 2023).

[8] Standard test method for airport pavement condition index surveys, in: Book of Standards Volume 04.03, October 2020, ASTM, 2020, pp. 0–55.

[9] J. Mccarthy, L. Fitzgerald, J. Mclaughlin, B. Mulry, D. O'brien, K. Dowling, Rural Flexible Roads Manual - Pavement Surface Condition Index, Vol 1 of 3, Department of Transport, Toursim and Sports, Dublin, Ireland, 2014. October 2014.

[10] SCANNER Surveys for Local Roads User Guide and Specification Volume 3 Advice to Local Authorities: Using SCANNER Survey Results. www.dft.gov.uk, 2011 (accessed May 23, 2023).

[11] Network Condition & Geography Statistics Branch, L.U.K, Department for Transport, Technical Note: Road Condition and Maintenance, London, 2021.

[12] Brian Mulry, Dr. Kieran Feighan, John McCarthy, Development and implementation of a simplified system for assessing the condition of Irish regional and local roads, in: 9th International Conference on Managing Pavement Assets, 2015, pp. 1–17.

[13] Brian Mulry, John McCarthy, A simplified system for assessing the condition of Irish regional and local roads, in: Civil Engineering Research in Ireland 2016, Dublin, 2016, pp. 1–7. https://ceri2016.exordo.com/files/papers/97/final_draft/097.pdf (accessed March 14, 2022).

[14] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, K. Murphy, Speed/accuracy trade-offs for modern convolutional object detectors, in: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. 2017-January, 2017, pp. 3296–3305, https://doi.org/10.1109/cvpr.2017.351.

[15] Z. Du, J. Yuan, F. Xiao, C. Hettiarachchi, Application of image technology on pavement distress detection: a review, Measurement 184 (2021), 109900, https://doi.org/10.1016/j.measurement.2021.109900.

[16] W. Cao, Q. Liu, Z. He, Review of pavement defect detection methods, IEEE Access 8 (2020) 14531–14544, https://doi.org/10.1109/ACCESS.2020.2966881.

[17] A. Ragnoli, M.R. De Blasiis, A. Di Benedetto, Pavement distress detection methods: a review, Infrastructures (Basel) (2018), https://doi.org/10.3390/infrastructures3040058.

[18] T.B.J. Coenen, A. Golroo, A review on automated pavement distress detection methods, Cogent Eng. 4 (2017), https://doi.org/10.1080/23311916.2017.1374822.

[19] C. Koch, K. Georgieva, V. Kasireddy, B. Akinci, P. Fieguth, A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure, Adv. Eng. Inform. 29 (2015) 196–210.

[20] Y. Hou, Q. Li, C. Zhang, G. Lu, Z. Ye, Y. Chen, L. Wang, D. Cao, The state-of-the-art review on applications of intrusive sensing, image processing techniques, and machine learning methods in pavement monitoring and analysis, Engineering 7 (2021) 845–856, https://doi.org/10.1016/j.eng.2020.07.030.

[21] N. Sholevar, A. Golroo, S.R. Esfahani, Machine learning techniques for pavement condition evaluation, Autom. Constr. 136 (2022), 104190, https://doi.org/10.1016/j.autcon.2022.104190.

[22] D. Arya, H. Maeda, S.K. Ghosh, D. Toshniwal, Y. Sekimoto, RDD2020: an annotated image dataset for automatic road damage detection using deep learning, Data Brief 36 (2021), 107133, https://doi.org/10.1016/j.dib.2021.107133.

[23] T. Rateke, K.A. Justen, V.F. Chiarella, A.C. Sobieranski, E. Comunello, A. Von Wangenheim, Passive vision region-based road detection: a literature review, ACM Comput. Surv. 52 (2019), https://doi.org/10.1145/3311951.

[24] S. Cano-Ortiz, P. Pascual-Muñoz, D. Castro-Fresno, Machine learning algorithms for monitoring pavement performance, Autom. Constr. 139 (2022), 104309, https://doi.org/10.1016/j.autcon.2022.104309.

[25] U.S. Department of Transportation, Federal Highway Administration. (2013, February). Practical Guide for Quality Management of Pavement Condition Data Collection. Retrieved April 25, 2023, from https://www.fhwa.dot.gov/pavement/management/qm/data_qm_guide.pdf.

[26] Lank, M. (2021). Road Quality Classification (Bachelor's thesis). Czech Technical University in Prague, Faculty of Information Technology. Retrieved from https://github.com/lenoch0d/road-quality-classification.

[27] K. Ma, M. Hoai, D. Samaras, Large-scale continual road inspection: visual infrastructure assessment in the wild, in: British Machine Vision Conference 2017, BMVC 2017, BMVA Press, 2017, https://doi.org/10.5244/C.31.151.

[28] New York City Department of Transportation. (2015). Street Pavement Rating [Dataset]. Retrieved from https://data.cityofnewyork.us/Transportation/Street-Pavement-Rating-Historical-/2cav-chmn/data#revert (Accessed April 25, 2023).

[29] City of Oakland. (2022, June). Pavement Condition Index (PCI) [Data source]. Retrieved April 25, 2023, from https://www.arcgis.com/apps/dashboards/5d844eacab5f40598fcd0e45376d785f.

[30] N. Sholevar, A. Golroo, S.R. Esfahani, Machine learning techniques for pavement condition evaluation, Autom. Constr. 136 (2022), 104190, https://doi.org/10.1016/j.autcon.2022.104190.

[31] J. Wang, Q. Meng, P. Shang, M. Saada, Road surface real-time detection based on Raspberry Pi and recurrent neural networks, Trans. Inst. Meas. Control. 43 (2021) 2540–2550, https://doi.org/10.1177/01423312211003372.

[32] B. Prasetya, Y.C.S. Poernomo, S. Winarto, R.K. Dewanta, F.M. Azhari, Mengurangi Laju Kerusakan Jalan dengan Menggunakan Metode RCI (road condition index) di

Kabupaten Madiun, Jurnal Manajemen Teknologi & Teknik Sipil. 4 (2021) 104–118, https://doi.org/10.30737/jurmateks.v4i1.1722.

[33] Y. Li, C. Liu, Y. Shen, J. Cao, S. Yu, Y. Du, RoadID: a dedicated deep convolutional neural network for multipavement distress detection, J. Transp. Eng. B: Pavements 147 (2021) 04021057, https://doi.org/10.1061/jpeodx.0000317.

[34] Y. Jiang, S. Han, Y. Bai, Development of a pavement evaluation tool using aerial imagery and deep learning, J. Transp. Eng. B: Pavements 147 (2021) 04021027, https://doi.org/10.1061/jpeodx.0000282.

[35] T. Nasiruddin Khilji, L. Lopes Amaral Loures, E. Rezazadeh Azar, Distress recognition in unpaved roads using unmanned aerial systems and deep learning segmentation, J. Comput. Civ. Eng. 35 (2021) 04020061, https://doi.org/10.1061/(asce)cp.1943-5487.0000952.

[36] I. Hashim Abbas, M. Qadir Ismael, Automated pavement distress detection using image processing techniques, Eng. Technol. Appl. Sci. Res. 11 (2021) 7702–7708, https://doi.org/10.48084/etasr.4450.

[37] T. Lee, Y. Yoon, C. Chun, S. Ryu, CNN-based road-surface crack detection model that responds to brightness changes, Electronics 10 (2021) 1402, https://doi.org/10.3390/electronics10121402.

[38] J. Menegazzo, A. von Wangenheim, Road surface type classification based on inertial sensors and machine learning: a comparison between classical and deep machine learning approaches for multi-contextual real-world scenarios, Computing 103 (2021) 2143–2170, https://doi.org/10.1007/S00607-021-00914-0.

[39] T. Rateke, A. von Wangenheim, Road surface detection and differentiation considering surface damages, Auton. Robot. 45 (2021) 299–312, https://doi.org/10.1007/S10514-020-09964-3.

[40] S. Zhou, W. Song, Crack segmentation through deep convolutional neural networks and heterogeneous image fusion, Autom. Constr. 125 (2021), https://doi.org/10.1016/j.autcon.2021.103605.

[41] D. Arya, H. Maeda, S.K. Ghosh, D. Toshniwal, A. Mraz, T. Kashiyama, Y. Sekimoto, Deep learning-based road damage detection and classification for multiple countries, Autom. Constr. 132 (2021), 103935, https://doi.org/10.1016/j.autcon.2021.103935.

[42] A. Issa, H. Samaneh, M. Ghanim, Predicting pavement condition index using artificial neural networks approach, Ain Shams Eng. J. (2021), https://doi.org/10.1016/j.asej.2021.04.033.

[43] Q. Chen, Y. Huang, H. Sun, W. Huang, Pavement crack detection using hessian structure propagation, Adv. Eng. Inform. 49 (2021), https://doi.org/10.1016/j.aei.2021.101303.

[44] R. Stricker, D. Aganian, M. Sesselmann, D. Seichter, M. Engelhardt, R. Spielhofer, M. Hahn, A. Hautz, K. Debes, H.-M. Gross, Road surface segmentation - pixel-perfect distress and object detection for road assessment, in: 2021 IEEE 17th International Conference on Automation Science and Engineering (CASE), 2021, pp. 1789–1796, https://doi.org/10.1109/case49439.2021.9551591.

[45] H. Majidifard, Y. Adu-Gyamfi, W.G. Buttlar, Deep machine learning approach to develop a new asphalt pavement condition index, Constr. Build. Mater. 247 (2020), 118513, https://doi.org/10.1016/j.conbuildmat.2020.118513.

[46] T. Rateke, K.A. Justen, A. Von Wangenheim, Road surface classification with images captured from low-cost camera-road traversing knowledge (RTK) dataset, in: Pdfs.Semanticscholar.Org 26, 2019, pp. 50–64, https://doi.org/10.22456/2175-2745.91522.

[47] A. Zhang, K.C.P. Wang, Y. Fei, Y. Liu, C. Chen, G. Yang, J.Q. Li, E. Yang, S. Qiu, Automated pixel-level pavement crack detection on 3D asphalt surfaces with a recurrent neural network, Comput. Aided Civ. Infrastruct. Eng. 34 (2019) 213–229, https://doi.org/10.1111/mice.12409.

[48] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, H. Omata, Road damage detection and classification using deep neural networks with smartphone images, Comput. Aided Civ. Infrastruct. Eng. 33 (2018) 1127–1141, https://doi.org/10.1111/mice.12387.

[49] J. Zhu, J. Zhong, T. Ma, X. Huang, W. Zhang, Y. Zhou, Pavement distress detection using convolutional neural networks with images captured via UAV, Autom. Constr. 133 (2022), 103991, https://doi.org/10.1016/j.autcon.2021.103991.

[50] D. Arya, H. Maeda, S.K. Ghosh, D. Toshniwal, A. Mraz, T. Kashiyama, Y. Sekimoto, Transfer Learning-based Road Damage Detection for Multiple Countries, 2020.

[51] A. Zhang, K.C.P.P. Wang, B. Li, E. Yang, X. Dai, Y. Peng, Y. Fei, Y. Liu, J.Q. Li, C. Chen, Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network, Comput. Aided Civ. Infrastruct. Eng. 32 (2017) 805–819, https://doi.org/10.1111/mice.12297.

[52] Y.A. Hsieh, Y. Tsai, Automated asphalt pavement raveling detection and classification using convolutional neural network and macrotexture analysis, Transp. Res. Rec. 2675 (2021) 984–994, https://doi.org/10.1177/03611981211005450.

[53] Y. Hou, Q. Li, C. Zhang, G. Lu, Z. Ye, Y. Chen, L. Wang, D. Cao, The state-of-the-art review on applications of intrusive sensing, image processing techniques, and machine learning methods in pavement monitoring and analysis, Engineering 7 (2021) 845–856, https://doi.org/10.1016/j.eng.2020.07.030.

[54] S. Hassan, D. O'sullivan, S. Mckeever, D. Power, R. Mcgowan, K. Feighan, Detecting patches on road pavement images acquired with 3D laser sensors using object detection and deep learning, in: International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISIGRAPP 2022 5, VISAPP, Scitepress, 2022, pp. 413–420, https://doi.org/10.5220/0010830000003124.

[55] A. Farhadi, J. Redmon, Yolov3: an incremental improvement, in: Computer Vision and Pattern Recognition, 2018.

[56] S. Mathavan, M.M. Rahman, M. Stonecliffe-Janes, K. Kamal, Pavement raveling detection and measurement from synchronized intensity and range images, Transp. Res. Rec. 2457 (2014) 3–11, https://doi.org/10.3141/2457-01.

[57] S. Ranjbar, F. Moghadas Nejad, H. Zakeri, Asphalt Pavement Bleeding Evaluation using Deep Learning and Wavelet Transform, Amirkabir Journal of Civil Engineering 53 (11) (2022) 1007–1010, https://doi.org/10.22060/ceej.2020.18292.6820.

[58] P.Y. Shinzato, T.C. Dos Santos, L.A. Rosero, D.A. Ridel, C.M. Massera, F. Alencar, M.P. Batista, A.Y. Hata, F.S. Osório, D.F. Wolf, CaRINA dataset: an emerging-country urban scenario benchmark for road detection systems, in: IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 2016, pp. 41–46, https://doi.org/10.1109/itsc.2016.7795529.

[59] J. Fritsch, T. Kuhnl, A. Geiger, A new performance measure and evaluation benchmark for road detection algorithms, in: IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 2013, pp. 1693–1700, https://doi.org/10.1109/itsc.2013.6728473.

[60] W.S. Qureshi, D. Power, M. Joseph, M. Brian, F. Kieran, O.S. Dympna, Learning pavement surface condition ratings through visual cues using a deep learning classification approach, in: 18th International Conference on Intelligent Computer Communication and Processing (ICCP 2022), Cluj-Napoca, Romania, 2022.

[61] OpenVINO Toolkit. (2022). road-segmentation-adas-0001. Retrieved December 2022, from https://docs.openvino.ai/2018_R5/_docs_Transportation_segmentation_curbs_release1_caffe_desc_road_segmentation_adas_0001.html.

[62] M. Tan, Q.V. Le, EfficientNetV2: Smaller Models and Faster Training. https://github.com/google/, 2021 (accessed October 3, 2022).

[63] Roadway by RoadBotics. https://roadway.demo.roadbotics.com/map/wPJQ8Zc82QxFHBbswpYs/?assessmentType=normal, 2022 (accessed February 3, 2022).

[64] Road Surface Inspection System | Global | Ricoh. https://www.ricoh.com/technology/tech/104_road_surface_monitoring, 2021 (accessed February 3, 2022).

[65] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, B. Schiele, D.A. R&D, T.U. Darmstadt, The Cityscapes Dataset for Semantic Urban Scene Understanding. www.cityscapes-dataset.net, 2016 (accessed March 2, 2022).

[66] CVAT: Image Annotation Tool. https://www.cvat.ai/, 2022 (accessed September 28, 2022).

[67] CVAT. (n.d.). CVAT: Image Annotation Tool. Retrieved 2023, from https://www.cvat.ai/.

[68] Golden. (n.d.). Mighty AI - Wiki. Retrieved 2023, from https://golden.com/wiki/Mighty_AI-YX9YV9V.

[69] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J.M. Alvarez, P. Luo, SegFormer: simple and efficient design for semantic segmentation with transformers, Adv. Neural Inf. Proces. Syst. 15 (2021) 12077–12090, https://doi.org/10.48550/arxiv.2105.15203.

[70] L.-C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking Atrous Convolution for Semantic Image Segmentation, 2017, https://doi.org/10.48550/arxiv.1706.05587.

[71] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2016-December, 2015, pp. 770–778, https://doi.org/10.48550/arxiv.1512.03385.

[72] Papers With Code. (n.d.). ResNet. Retrieved from https://paperswithcode.com/lib/timm/resnet (accessed October 7, 2022).

[73] Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid Scene Parsing Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 6230-6239). DOI: 10.1109/CVPR.2017.660.

[74] Sarkar, D. (2022). A Comprehensive Hands-on Guide to Transfer Learning with Real-World Applications in Deep Learning [Blog post]. Towards Data Science. Retrieved from https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a (accessed October 12, 2022).

[75] S.P.G. Jasil, V. Ulagamuthalvi, Deep learning architecture using transfer learning for classification of skin lesions, J. Ambient. Intell. Humaniz. Comput. 2021 (2021) 1–8, https://doi.org/10.1007/S12652-021-03062-7.

[76] H.C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R. M. Summers, Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, IEEE Trans. Med. Imaging 35 (2016) 1285–1298, https://doi.org/10.1109/tmi.2016.2528162.

[77] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, S. Xie, F.A. Research, A ConvNet for the 2020s, in: IEEE Computer Vision and Pattern Recognition Conference, 2022. https://arxiv.org/abs/2201.03545v2 (accessed October 6, 2022).

[78] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, F. Wei, B. Guo, Swin transformer V2: scaling up capacity and resolution, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Institute of Electrical and Electronics Engineers (IEEE), 2022, pp. 11999–12009, https://doi.org/10.1109/cvpr52688.2022.01170.

[79] Hendrycks, D., & Gimpel, K. (2023). Gaussian Error Linear Units (GELUs). arXiv preprint arXiv:1606.08415v5 [cs.LG].DOI: 10.48550/arXiv.1606.08415.

[80] OpenMMLab. (2022). mmsegmentation: OpenMMLab Semantic Segmentation Toolbox and Benchmark [GitHub repository]. Retrieved from https://github.com/open-mmlab/mmsegmentation (accessed October 25, 2022).

[81] OpenMMLab. (2022). Retrieved 2023, from https://openmmlab.com/OpenMMLab. 2023 https://openmmlab.com/ (accessed October 25, 2022).

[82] Rwightman. (2022). pytorch-image-models: PyTorch Image Models, Scripts, Pretrained Weights – ResNet, ResNeXT, EfficientNet, EfficientNetV2, NFNet, Vision Transformer, MixNet, MobileNet-V3/V2, RegNet, DPN, CSPNet, and More [GitHub repository]. Retrieved from https://github.com/rwightman/pytorch-image-models#introduction (accessed October 25, 2022).

[83] Weights & Biases. (n.d.). Weights & Biases – Developer Tools for ML. Retrieved 2023, from https://wandb.ai/site.

[84] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, ImageNet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (2015) 211–252, https://doi.org/10.1007/s11263-015-0816-y/figures/16.

[85] T. Ridnik, E. Ben-Baruch, A. Noy, L. Zelnik-Manor, ImageNet-21K Pretraining for the Masses, 2021, pp. 1–20. http://arxiv.org/abs/2104.10972.

[86] Model Zoo Summary — MMClassification 0.24.0 Documentation. (n.d.). Retrieved from https://mmclassification.readthedocs.io/en/master/modelzoo_statistics.html (accessed October 12, 2022).

[87] Rwightman. (n.d.). PyTorch Image Models, Scripts, Pretrained Weights – ResNet, ResNeXT, EfficientNet, EfficientNetV2, NFNet, Vision Transformer, MixNet, MobileNet-V3/V2, RegNet, DPN, CSPNet, and More. GitHub. Retrieved from https://github.com/rwightman/pytorch-image-models (accessed October 12, 2022).

[88] PyTorch. (n.d.). Model Zoo — PyTorch/Serve Master Documentation. Retrieved from https://pytorch.org/serve/model_zoo.html (accessed October 12, 2022).

[89] TensorFlow. (n.d.). Models & Datasets. Retrieved from https://www.tensorflow.org/resources/models-datasets (accessed October 12, 2022).

[90] Model Zoo. (n.d.). Model Zoo - Deep Learning Code and Pretrained Models for Transfer Learning, Educational Purposes, and More. Retrieved 2023, from https://modelzoo.co/.

[91] Papers With Code. (n.d.). Image Classification. Retrieved from https://paperswithcode.com/task/image-classification (accessed October 10, 2022).

[92] C.K. Ryali, D.J. Schwab, A.S. Morcos, Characterizing and Improving the Robustness of Self-Supervised Learning through Background Augmentations, 2021, https://doi.org/10.48550/arxiv.2103.12719.

[93] J.L. Fleiss, J. Cohen, B.S. Everitt, Large sample standard errors of kappa and weighted kappa, Psychol. Bull. 72 (1969) 323–327, https://doi.org/10.1037/H0028106.

[94] J.L. Fleiss, J. Cohen, The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability, Educ. Psychol. Meas. 33 (1973) 613–619, https://doi.org/10.1177/001316447303300309/asset/001316447303300309.fp.png_v03.

[95] J.L. Fleiss, J. Cohen, The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability, Educ. Psychol. Meas. 33 (1973) 613–619, https://doi.org/10.1177/001316447303300309.

[96] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J.M. Alvarez, P. Luo, SegFormer: simple and efficient design for semantic segmentation with transformers, Adv. Neural Inf. Proces. Syst. 15 (2021) 12077–12090. https://github.com/open-mmlab/mmsegmentation (accessed November 7, 2022).

[97] MMSegmentation Contributors. (2020). MMSegmentation: OpenMMLab Semantic Segmentation Toolbox and Benchmark. Retrieved October 2022, from https://github.com/open-mmlab/mmsegmentation.