Instance Segmentation for Car Damage Detection with Mask-RCNN

Rahul Rudra [1] , Praloy Sarker [2] , Farhan Hai Khan [3], Amartya Guha [4] and Romit Maity[5]

Institute of Engineering and Management[1-5], rahulrudra245@gmail.com, praloy40@gmail.com, njrfarhandasilva10@gmail.com, amartya.aa99@gmail.com, romitrmaity@gmail.com

Abstract

Detection of damage on cars is a task achievable by an image-based recognition method. This method of detection and calibration of exterior damage on a car would prove to be extremely useful for car insurers, car rentals, and car-repair services. Car damage reporting and penalty calculations have always been a challenging issue for licence companies and used car-selling companies. In this context, we put forward the application of deep learning and neural networks for building an architectural model of Mask R-CNN which can quantify the damaged surface area of the car. We use a deep feature prediction network for accurate instance segmentation of the damaged car parts. This study is an extension of business technologies to detect and quantify car scratches to address the problems faced by the used car industry and car rental companies. It will support businesses in eliminating middlemen and paving the way for a more objective system of pricing and insurance in the vehicle dealership market. For the proposed research a customised novel dataset was used consisting of images of damaged vehicles from multiple scenarios with precise labelling and annotation for image segmentation, and the Mask-RCNN model achieves a suitably high accuracy of over 93%.

Keywords. Mask-R CNN; Instance Segmentation; Image Recognition, Car Damage Detection, Deep Neural Networks, Object Detection, Machine Learning, Computer Vision, Deep Learning, Artificial Intelligence, Industry 5.0, Industrial Automation.

1. Introduction

For optimal damage detection for optimal damage detection and assessment, deep learning has been the state-of-the-art approach. For this study, the Mask-RCNN model detects the classified portion of the item and recognizes its location with severe damage. In the era of digitalization, machine vision has been introduced for a better solution by applying Automated Visual Inspection (AVI) [1] systems to have a quality output for eradication of defective products. we should follow machine-assembled auto inspection algorithms in industries to facilitate industrial automation.

For example, bright glass is one of them used on the outer sides and the threshold of the building in the wave of ultra-modern architecture [2]. The image-based surface scratches are created in the manufacturing process lines that lead to making weak materials. To have the uncorrupted products an engineer should be appointed for scratches and region-based Mask R-CNN ([1][2]) close to the surface of the bright glass items. Moreover, in the 21st-century of evolution, vehicular travel has increased exponentially, leading to hit-and-run cases and vehicle crashes. These haphazard road accidents can be detected for vital timed information to hospital authorities and can also be useful to insurance industries for vehicle rentals. They can further use the damage assessment to create reports for damage cost estimation with the vehicle model, base price, insurance premium values, and more to get a correct cost estimation approach. [4] Automated microscope camera scanning systems are needed to collect the image with scratches for building training and validation datasets for damage classification, this is essential for damage detection as manual humane approaches are often time taking and biased towards opinions and sometimes fueled by corrupted middlemen.

The prevailing market of the automobile industry directly leads to a substantial increase in the number of severe car accidents. With reference to the context, here we quantify deep-learning-based neural network architectural backbones with Mask-RCNN, viz., VGG-16 and VGG-19, for the inspection of car damage detection and its real-life assessment datasets [3]. With the increase of cars on the roads, the figures for car accidents have drastically increased. As a result, car damage is also increasing at an exponential rate [4]. So, it has become very important to identify car damages. We explore deep learning techniques for this purpose.

The proposed algorithm detects the damaged part of the car followed by its location and the intensity of the damage. An application of transfer learning for pre-trained VGG models over the Microsoft COCO (Common Objects in Context) [4] is being used to decrease the error rate of our proposed model. However, the characteristics classification of such phenomena is challenging due to scarcely visible damages and correlated inter-class mixing of damages with the vehicle body.

Many traditional computational-based machine-learning techniques were used to find out an automatic way out of the vehicle damage problem. Vehicle scratch detection is obtained by rendering the 3D CAD model of vehicles over the damaged dataset. Several attempts were made to come up with a sustainable solution with the help of crown detection and satellite image processing. We used state-of-the-art object detection methods using CNN to train the damage detection model with our custom-labeled dataset [5]. Upon comparison with traditional target methods the target detection Mask-RCNN model architecture proves to be effective in case of small target detection [4], which is widely used in several domains - agriculture, construction, healthcare, and other fields.

The use of machine vision in case of identifying defects of the industrial finished goods is considered as surface defect inspection. These defects include but are not limited to scratches, pits, aberration, dust-sport, and stains. Manual inspection methods have been the traditional approaches and nowadays are not sufficient to achieve these purposes, on the other hand, the modern method of advanced computer vision is the only key method to identify the defects on the surface of the finished goods. Applying the automated visual inspection (AVI) [6] system we can easily recognize defective goods and remove them from the production scheme. Automated Visual Inspection (AVI) is itself the practised-hand inspector to solve all problems relating to surface defects inspection. The world's leading economic giants like the USA, India, Japan, Europe, and China have created a vast market all over the richest countries to perform their achievements using machine-based technology that was used for advanced deep learning [1, 20]. The system has also been upgraded such that it has a continued process for auto-detection and auto-removal as well. Instance segmentation is the science of image classification where each pixel of the image is assigned a class label. In other words, instance segmentation is the task of detecting and delineating each distinct object of interest appearing in an image. Mask R-CNN is the current state-of-the-art technology [8] for highly accurate mask detection for RoIs (Region of Interest). In this paper, we train the M-RCNN model to train and detect effective damage areas in an image/video [7]. This paper also presents a business extension of existing technologies to detect car scratches and quantify damages, in order to tackle the problems faced by the used car industry, traffic control system[17] and car rental companies for the automation of penalties that occurred due to these accidents.

Previous attempts at claims processing automation have been done using various approaches which have provided the bounding boxes for the scratches. But, due to the positional properties of vehicular damages, these detected bounding boxes do not cover the exact portions of the detected scratches and rather build up rectangles around them, which isn't feasible for calculating accurate damage costs. This research maps the damages to a multipoint feature polygon with various dynamic shapes rather than static ones.

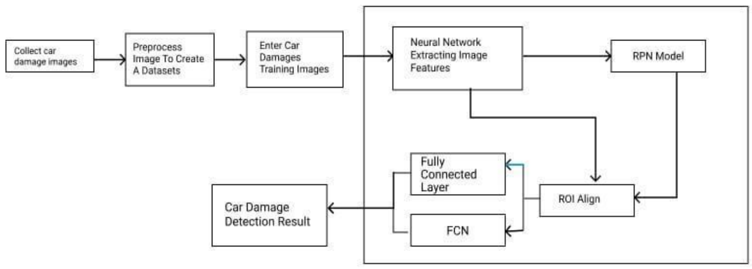


Fig.1: Car Damage Detection-Segmentation Framework

The image of the damaged part of the car is selected and collected according to the need and the data is marked with the VGG Image Annotator (VIA) tool from the Visual Geometry Group and converted into the widely used JavaScript Object Notation (JSON)[19] format which is further divided into a training set and test set with an 80%: 20% (762:191 images) split ratio with a total of 953 images obtained from the internet and labelled by the team. The algorithm here works in two ways. Firstly, the region proposal network scans and generates the Region of Interest (RoI). The second stage processes the proposal from the network and generates a bounding box and mask based on the class confidence. The car damage detection and segmentation system based upon this piece of research are presented in Fig.1.

1.1. Algorithmic Architectural Workflow:

(1) Pass an image for processing to a pre-trained ResNet50 + FPN (Feature Pyramid Network) network model to extract features and find compatible feature mappings.

(2) The feature map acquires a large number of candidate frames via RPN and uses a SoftMax separator to create dual front and rear editing, using these frames to determine a more accurate position of the candidate's frames. The information and filtering part of the ROI with non-high-pressure compression is performed next.

(3) Feature map and final ROI remaining in the RoIAlign layer, so that each ROI produces a feature map of a fixed size.

(4) Finally, flow flows through two branches, one branch entering a fully integrated layer of layout and deceleration and the other into a full network of pixel separation variables.

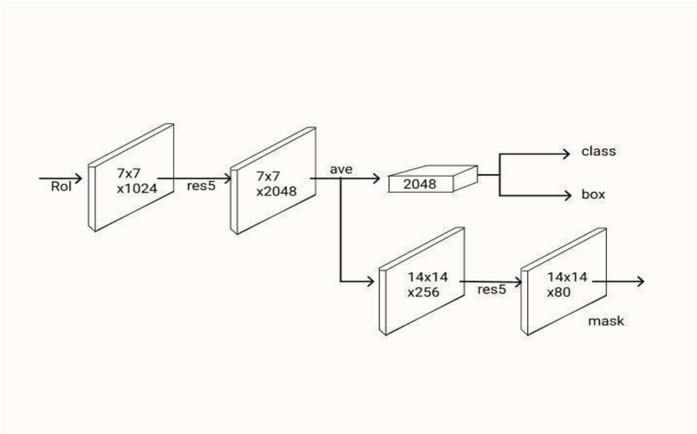


Fig 2: Mask Region-based CNN: M-RCNN Model built on the Faster-RCNN model architecture and used the Res-Net C4 map features

2. Machine Learning Pipeline

The ML Pipeline consists of four categories: Data Collection and Annotation, Network Training, Model Validation, and Model Deployment. Firstly, the data is collected for training the neural network, then we label the data using VGG Image Annotator (VIA) [14]) and the network is trained using TensorFlow and Keras Framework on Python with NVIDIA GPU Accelerations[18]. The source implementation is modified from Facebook AI and Research Team's Matterport Implementation . We then compute the model accuracy and finally, we deploy the model using Flask.

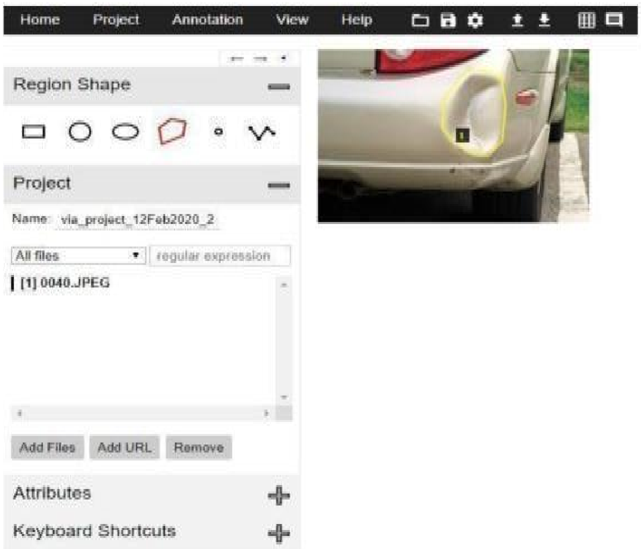
3. Detailed Methodology

3.1. Dataset Collection and Annotation Creation

As there is no publicly available data set for car damages by category, we have created our own novel dataset that includes images for different types of car damage. We consider seven types of damage often referred to as bump rot, door hammer, glass shatter, broken headlamp, broken tail lamp, scratch and crush in our image set for adding variance to the data. In addition, we also collected photos of the un-damaged cars for better training and to avoid false positives via overfitting. Also, two sets of classes were divided and labeled, which were the "background" and "scratch" to perform binary pixel-wise classification.

The proposed architecture constitutes deep learning with neural networks to mask and detect damages discovered on a vehicle's body. We use the state-of-the-art Mask-RCNN model to train our own custom model and get 93% accurate results to our predictions.

For this manuscript, we used the dataset from Kaggle [12, 21]. After performing some basic data cleaning, we annotated the image using the official VGG Image Annotator (VIA Tool) [14]. This annotator gives us our JSON (JavaScript Object Notation) file, which we use later for training and ground truth (GT) comparison.



3.2. Data Annotation

After collecting the images, annotating them is the next step. The annotations will clearly demarcate the area of damage in each training image, which will help the model learn when training. The machine learning model requires the images in the training dataset to be annotated, which is to have the region of damage in an image identified and the boundary of the damaged area marked accurately. A sample example of an annotation from the used VGG Image Annotator (VIA), is shown in Fig.3. The annotations of the images are then stored in JSON format in the directory of the dataset and will be used while training.

3.3. Import dataset and Visualization

With the dataset and the annotations ready, the next step involves Data Wrangling and Data Visualization. Our version of Mask R-CNN is customised and built on top of the Matterport Mask RCNN GitHub repository from the Facebook AI and Research Team (FAIR). Images and annotations from the dataset will be loaded using a custom Data Loader class function with a configuration for training and inferences. Visualisation of Input (Images and Masks) can give us a view of how the mask would look when placed on the image and hence how the model would see it, also the variation in the data is important to prevent model overfitting.

In order to handle bounding boxes more consistently, it is better to compute the coordinates from the damage mask for the car images, instead of using these coordinates as a part of the input dataset.

3.4. Visualisation of Input (Images and Masks)

In Fig 4, we have taken five input images and have plotted the image along with their damage masks. Visualisations approve whether the annotation was done correctly or not.

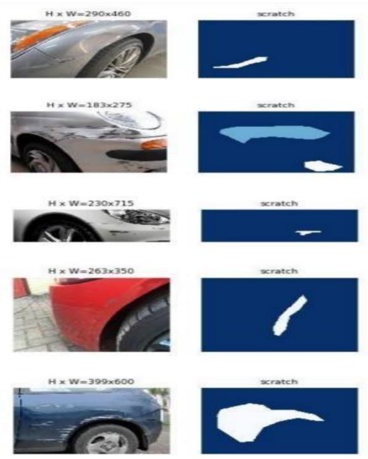


Fig.4 Plot of a subset from the full training dataset with their damage mask, image height included.

3.5. Bounding Box (BB) with annotated damage mask for typical car images

With the bounding box being computed from the damage mask and not passed as input, we can see that the coordinates of the bounding box give us a better estimation of the damaged area. This is shown in Fig 5, where it can be seen that the bounding box completely encloses the damage mask and hence gives us an idea that computing the coordinates by using this method will make our methodology robust and consistent over multiple images.

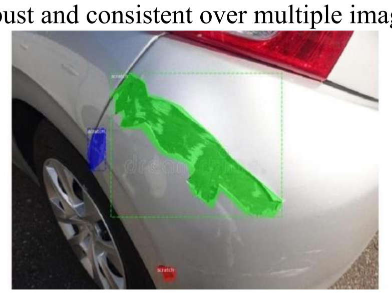


Fig.5 Damage mask on the car image with bounding boxes

3.6. Training model

Data was converted from VIA Image Annotator to valid COCO Annotation formats. Model heads were trained for 50 epochs and Fine Tuning of the model was performed for another 50 epochs. For training the model heads, the other weights are frozen and only the final layers are trained for the custom classes for faster convergence rates, whereas fine-tuning performs model optimization by modifying overall weights with slight adjustments. Both Dice Loss and RoI loss were chosen as the loss metrics for the image segmentation task. The model was trained with an initial learning rate of 0.01. The model was finally evaluated on the Intersection Over Union Score (IoU) for better accuracy and faster convergence, pre-trained weights from the COCO dataset were used, instead of training the model from scratch using the Adam Optimizer, and ReduceLRonPlateau with patience of 15 epochs.

Data was split in an 80-20 ratio for training and testing. For instance, segmentation, the Mask-RCNN model was used. The losses used are explained below:

Dice Loss:

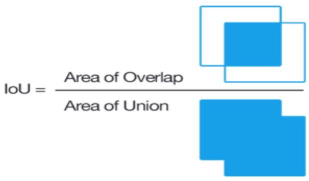
The Dice coefficient (DSC) is a measure of how closely two sets overlap. For instance, if two sets A and B perfectly overlap, DSC reaches its maximum value of 1. Otherwise, DSC begins to decline, eventually reaching 0 if the two sets do not overlap at all as referred in the next image [i].

$$D = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2}$$

Intersection over Union:

IoU is the area of overlap between the predicted segmentation and the ground truth divided by the area of union between the predicted segmentation and the ground truth, as shown on the image[ii] to the left.

3.7. Model Validation



The model weights are inspected by plotting the variance in the weights. Link last training checkpoint for model for validation. This step performs a sanity check if your weights and biases are properly distributed. Perform a sanity check if weights and biases are properly distributed.

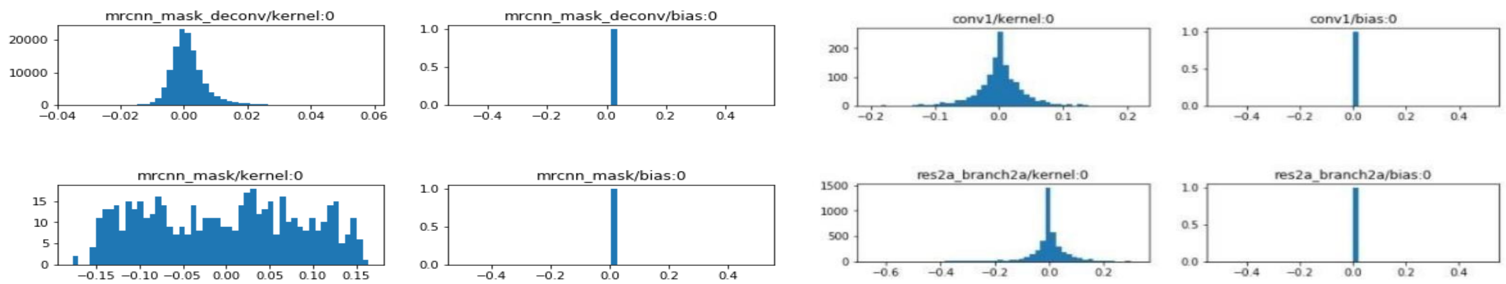
Model	Precision	Recall	Accuracy without augmentation	Accuracy with augmentation
[22] VGG16	0.94	0.94	94.56	94.56
[22] VGG19	0.91	0.91	94.57	95.22
[23] Inception	0.6175	0.5671	71.82	71.50
[23] Alexnet	0.6142	0.5809	70.85	73.91
[23] Resnet	0.8438	0.8110	88.24	87.92
[24] VGG19	0.9220	0.9080	93.20	94.90
[24] Resnet50	0.928	0.922	89.58	90.26

Table 1

The following Table 1 we can understand how precise a model is out of those predicted positives, and how many of them are actual positives. Precision is a good measure to determine accuracy only when the cost of negative is high. Recall actually measures how many of the actual positives a model captures by labelling it as positive. Applying the same understanding, we know that recall will be the model metric that we use to select our best model when a high value is associated with a positive. The model accuracy depends on precision and recall value. Accuracy can tell immediately if a model is being trained properly and how it can work in general. All performance accuracy of VGG19 is better than VGG16 even if its values of matrices are not larger than VGG16. Inception deep convoluted architecture, also known as GoogleNet, was developed by Google. The model consists of a basic unit and 'inception cell' where the performance of a series of convolutions on different scales and subsequently combine the result. Alexnet is one of the deepest convents designed to deal with complex scenes and ImageNet data classification work. The network was very similar to LeNet, but deeper, large and characteristic convoluted layers piled on top of each other. Per the low overfitting, AlexNet uses another technique called dropout. Unlike traditional sequential network architectures such as AlexNet, Overfit, and VGG, a form instead of ResNet of ''foreign architecture'' which relies on micro-architecture modules.

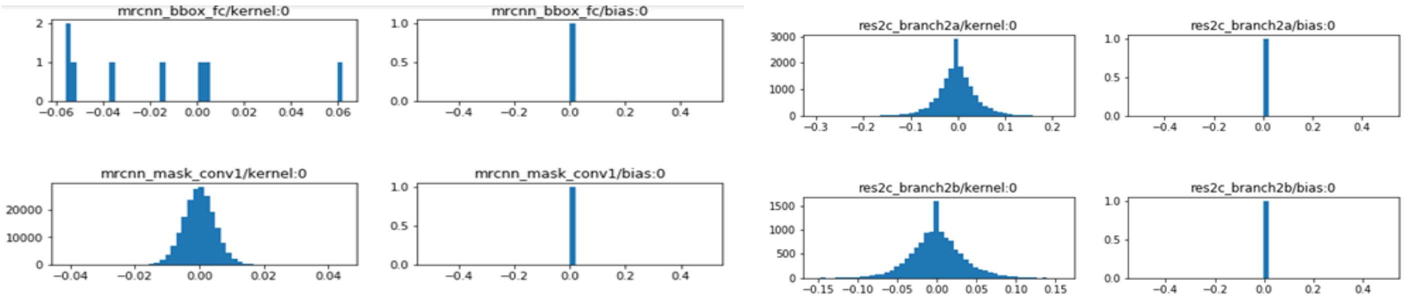
3.8. Prediction

Inspect model by performing prediction on test and validation to test the accuracy. Prediction on test image.



Damage Localization

With the same procedure, the authors have trained the dataset in such a way as to localize the damaged part. For each pixel in the test image,100 x 100 size is cropped all around then 224x224 predicts the background of the class. Damage is predicted by consideration of the fact if the particle exceeds a certain limit.



4. Business Implementation and Road Ahead

Car rental companies can install such proposed infrastructure architecture with the help of high-resolution cameras at different locations at particular angles for capturing standardized images of different body parts of the car, which can detect all possible exterior damages in the car. This software application can be advanced as a Mobile API which can further ease the verification process after detection and mask filter feature for the damaged car. This system of approach will be helpful for the rental evaluators, and resolution personnel in calibrating the relative damage area of the car. The best thing here is the users do not have to collect too many images or annotations as the architectural model training starts from trained weights('coco'), hence the time complexity is very less. This software technology can further be extended into detection systems of different types of visible car damages and its respective faults. The demand for improved customer service and improved experience is now so universal that customers are spoiled for choice. Therefore, it is not at all surprising. From booking a rental car to re-submitting a ticket, customers want a seamless experience at every stage. One can only make it faster and more responsive through automated technology.In case of damage here should be considered some points they are the location of damages, patterns of damages & types of damages to the car. The detection system of the damage must be developed for better information for the customers or users relating to rental cars. There is a higher need for automation to restore confidence in the car rental industry. Fortunately, the automation of car inspections and car damage assessments with AI makes fast, accurate, and data-driven deductions possible. Most importantly, these AI inspections are backed by documentary evidence, which improves reliability.

5. Conclusion Further Improvements

Our model performs quite well on random images from the Internet and generalises well for the scratches. This summarises that the model is trained well and has learned the scratch detection algorithm efficiently. Since our dataset (2400 images) is much smaller than the big AI datasets, viz., COCO (1,20,000 images), we do not expect to get better results from our model using the same approach. However, pre-segmentation of vehicles and then damage detection could be an improvement in the algorithm. The proposed algorithm relies on two steps: a pre-processing step, performed during the scanning process, and a post-processing step, where, using the information from the pre-processing step, defects can be detected with ease. And considering the analysis of the results, a claim can be made that the benefit from training on larger datasets will only magnify the performance of the model. Currently more and more research is being done towards creating deeper and different convolutional networks, which enables it to be used to address an increasing number of scenarios. As such, future work could focus on the application of similar models for the detection of imperfections on objects in other domains.

6. Acknowledgement

We'd like to thank Mr. Arjun Dutta, Product Architect at IQApex Labs for sponsoring us with the project and other necessary information regarding the design and development of the algorithm and conceptualizing the whole manuscript and our institution and professors for the opportunity and for guiding the way.

7. References

[1] Recent advances in surface defect inspection of industrial products using deep learning techniques. The International Journal of Advanced Manufacturing Technology (2021) 113:35–58. https://doi.org/10.1007/s00170-021-06592-8

[2] Image-based surface scratch detection on architectural glass panels using a deep learning approach. https://doi.org/10.1016/j.conbuildmat.2021.122717

[3] Phyu Mar Kyu and Kuntpong Woraratpanya. S.: Car Damage Detection and Classification. In: Proceedings of International Conference on Advances in Information Technology (IAIT2020), July 1-3, 2020, Bangkok, Thailand. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3406601.3406651

[4] Ranjodh Singh; Meghna P. Ayyar; Tata Venkata Sri Pavan; Sandeep Gosain; Rajiv Ratn Shah. S .: Automating Car Insurance Claims Using Deep Learning Techniques. In: 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM). https://10.1109/BigMM.2019.00-25

[5] Hiroya Maeda, Yoshihide Sekimoto, Toshikazu Seto, Takehiro Kashiyama, Hir Omata. S.: Road Damage Detection and Classification Using Deep Neural Networks with Smartphone Images, First published: 30 June 2018 https://doi.org/10.1111/mice.12387

[6] A generalized model for scratch detection. IEEE Transactions on Image Processin (Volume: 13, Issue: 1, Jan. 2004). DOI: 10.1109/TIP.2003.817231

[7] Car Damage Detection and Analysis Using Deep Learning Algorithms for Automotive. International Journal of Scientific Research Engineering Trends Volume 5, Issue 6, Nov-Dec-2019, ISSN (Online): 2395-566X

[8] Assessing Car Damage with Convolutional Neural Networks. 2021 International Conference on Communication information and Computing Technology(ICCICT) | 978-1-6654-0430-3/21/31.002021*IEEE* | *DOI:*10.1109/*ICCICT*50803.2021.9510069

[9] 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA). DOI: 10.1109/ICMLA.2017.0-179

[10] Hermans, L; Van der Auweraer, H; Mevel, L. S.: Health monitoring and Detection of a fatigue problem of a sports car, In: IMAC - PROCEEDINGS OF THE 17TH INTER- NATIONAL MODAL ANALYSIS CONFERENCE, VOLS I AND II, 3727 Pages: 42 - 48, lirias.kuleuven.be/875533?limo=0

[11] Blogs - Towards Data Science (Medium) Analytics Vidhya, Sourish Dey,Priya Dwivedi, @AnalyticsVidhya , analyticsvidhya.com/blog/2018/07/building-mask r-cnn-model-detectingdamagecars-python/,@Medium, towardsdatascience.com/cnn-application-detecting-car-exterior damage-full implementable-code-1b205e3cb48c

[12] Car damage detection - Kaggle Dataset, Anuj Shah (@anujms), Dataset- Car damage detection, kaggle.com/anujms/car-damage-detection

[13] An Expert System to Detect Car Damage by using CART Method. Prof. Dr. IKetut Gede Darma Putra, S.Kom., M.T., I KETUT GEDE DARMA PUTRA and Ni Made Ika Marini Mandenni, ST.,M.Kom, NI MADE IKA MARINI MANDENNI (2014) An Expert System to Detect Car Damage by using CART Method. Journal of Theoretical and Applied Information Technology (JATIT), 66 (3). ISSN 1992-8645. https://erepo.unud.ac.id/id/eprint/9794

[14] Abhisek Dutta; Andrew Zisserman, S.: The VIA Annotation Software for Images, Audio and Video, In: MM '19: Proceedings of the 27th ACM International Conference on Multimedia, pp. 2276–2279. https://doi.org/10.1145/3343031.3350535

[15] Automatic Paint Defect Detection and Classification of Car Body. Parisa Kamani, Elaheh Noursadeghi, Ahmad Afshar, Farzad Towhidkhah. 2011 7th Iranian Conference on Machine Vision and Image Processing. DOI:10.1109/IranianMVIP.2011.6121575

[16] Scratch Detection in Cars Using a Convolutional Neural Network by Means of Transfer Learning. International Journal of Applied Engineering Research ISSN

0973-4562 Volume 13, Number 16 (2018) pp. 12976-12982 © Research India Publications. http://www.ripublication.com

[17] A. Dutta et al., "Intelligent Traffic Control System: Towards Smart City," 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2019, pp. 1124-1129, doi:10.1109/IEMCON.2019.8936188.

[18] Mask R-CNN Keras Implementation (@matterport), Waleed Abdulla, Mask RCNN Implementation, github.com/matterport/Mask*RCNN*

[19] E. Rathje and M. Crawford, S.: Using high resolution satellite imagery to detect damage from the 2003 northern Algeria earthquake, In: 13th World Conference on Earthquake Engineering, August 2004, pp. 1–6.

[20] Ayantika Dey and Darothi Baidhya, *Development of Intelligent Car Parking System*, vol. 1, no. 4, 2020, ISSN 2689-484X.

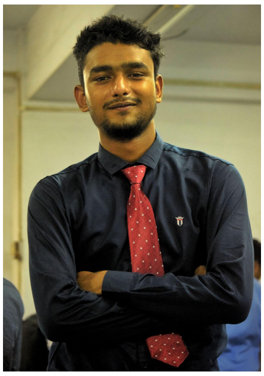
[21] A.Dutta et al., Farhan Hai Khan,COVID-19: Detailed Analytics & Predictive Modelling using Deep Learning, DOI: 10.32628/IJSRST207517

Biographies

Rahul Rudra is currently pursuing bachelor's degree in computer science and engineering from the Institute of Engineering and Management, Kolkata, under MAKAUT. His interests and research areas include cyber security and machine learning

Romit Maity obtained his higher secondary education from DAV Public School and is currently pursuing a bachelor's degree in computer science and engineering from the Institute of Engineering and Management. His interests and research areas include deep learning, artificial intelligence and bioengineering.

Praloy Sarker obtained his higher secondary education from Ispahani Public School and College, Cumilla, Bangladesh and is currently pursuing a bachelor's degree in Information Technology from the Institute of Engineering and Management. His interests and research areas include deep learning and artificial intelligence.



Farhan Hai Khan is a Software Engineer and an Independent Researcher with experience working wth competitive teams and emerging startups. His background in Artificial Intelligence, Machine Learning, Computer Vision & Deep Learning inform his technical and scientific approaches.He received the Director's Award for the Best Scientific Mind while pursueing Electrical Engineering from IEM Kolkata. He has also published research papers, journals & articles in reputed press such as IEEE, Springer, Taylor & Francis etc.