IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.10527

Analysis and design of an automated insurance claim platform with fraud detection

Andria Maria Ajai[1], Amala Abraham[2], Bodhisha Thomas[3], Saurav E S[4], Cerene Mariam Abraham[5]

Student, Dept.Of Computer Science, Muthoot Institute of Technology & Science, Ernakulam, India[1,2,3,4]

Assistant Professor, Dept.Of Computer Science, Muthoot Institute of Technology & Science, Ernakulam, India[5]

Abstract: Image and video-based vehicle insurance claim processing is an important area with a large scope for automation in the insurance sector. At present, the insurance claim process and validation are done manually resulting in a time-consuming process with less scalability and more prone to error. Especially during the pandemic times, manual inspection is a difficult process and claim amounts primarily rely on the type of damage and damaged part of the car, so rise the need for an automated system for the whole process of car insurance claim as which can efficiently classify and detect damage and helps to minimize the claim leakage. Also, there is a chance of faking car damage images using image forgery or deepfake generation techniques. To analyze and design an automated vehicle insurance claim platform that can perform car damage detection and classification along with image forgery & deepfake detection.

Keywords: Transfer learning, YOLO, image copy-move forgery detection (CMFD), speeded-up robust feature (SURF), polar complex exponential transform (PCET), Deepfakes, LSTM

I. INTRODUCTION

Today, in the vehicle insurance industry, claims leakage occurs at a higher rate which results in wastage of a lot of money. The difference between the actual insurance claim payment made and the sum which should have been paid if all the industry-leading practices were applied is known as the claim leakage. To reduce such effects visual inspection and validation have been used. However, these techniques result in delays in claim processing. There have been efforts by too few start-ups to mitigate claim processing time. An automated vehicle insurance claim platform is the need of the hour.

As per the proposed solution, the client uploads clear images and videos of the insured vehicle along with the Licence and RC book to the platform where it automatically verifies the images and videos. There may be certain cases where the user can upload fake images to claim insurance. In our proposed system, we are introducing a new feature that helps to detect forged images/deepfake images given by the user. Convolutional Neural Network (CNN) based methods are employed for the classification of car damage types. Based on the classification, the machine learning model will generate a detailed report of the damage which includes information about the damaged parts and their severity. A cost estimation model is also employed which estimates the overall cost of the damage, based on which the insurance claim can proceed with ease. Fig. 1 shows the architectural diagram of the proposed solution.

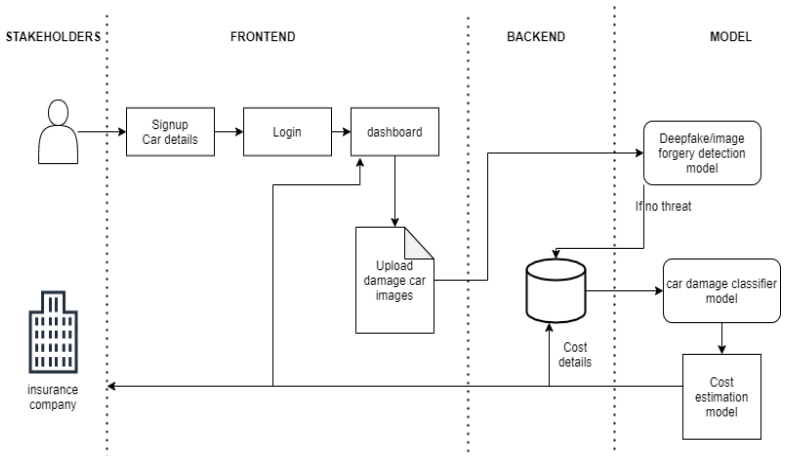


Fig. 1. An automated vehicle insurance claim platform with detection of Deepfakes

This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.10527

II. DAMAGE DETECTION AND CLASSIFICATION

A. Transfer Learning

In transfer learning, the cognizance of an already trained machine learning model is applied to a different but cognate quandary. The general conception is to utilize the erudition a model has learned from a task with an abundance of available labeled training data in an incipient task that doesn't have much data. In lieu of starting the cognition process from scratch, we commence with patterns learned from solving a cognate task. Convolutional neural networks conventionally endeavor to detect edges in the earlier layers, shapes in the middle layer, and some task-concrete features in the later layers. In transfer learning, the early and middle layers are utilized and we only retrain the latter layers. It leverages the labeled data of the task it was initially trained on[4].

There are different approaches when we consider applying transfer learning. One, Training a model to reuse it. Two, Utilizing a pre-trained model. Three, Feature extraction. Here when we consider conveyance damage relegation we are going to utilize the second approach which is to utilize a pre-trained model. There are a plethora of models out there that can be used in transfer learning. How many layers to reuse and how many to retrain depends on the quandary. Keras, for example, provides nine pre-trained models that can be utilized for transfer learning, prognostication, feature extraction, and fine-tuning.

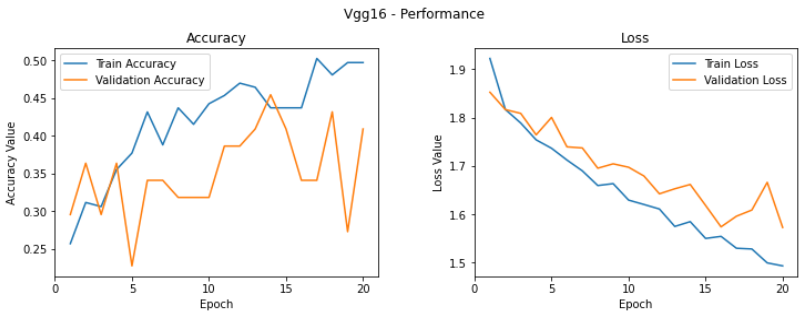
VGG16. VGG16 is a convolutional neural network model put forward by K. Simonyan and A. Zisserman from the University of Oxford. The model was able to achieve a 92.7% top-5 test accuracy in the ImageNet dataset, which is a dataset of over 14 million images belonging to mainly 1000 classes. It was one of the famous models that have been submitted to ILSVRC-2014.

VGG19. VGG19 is a convolutional neural network model put forward by K. Simonyan and A. Zisserman from the University of Oxford. VGG19 is a variant of the VGG model which actually consists of 19 layers (16 convolution layers, 5 MaxPool layers, 3 Fully connected layers, and 1 SoftMax layer) whereas VGG16 consists of only 16 layers. It has been used just as a good classification architecture for many other datasets and as the authors have made the models available to the public, so they can be used as it is or with modification for other similar tasks too[5].

AlexNet. AlexNet was designed by Alex Krizhevsky, is one of the deep ConvNets designed to deal with complex scene classification task on Imagenet data. The architecture of AlexNet consists of eight layers: five convolutional layers and three fully-connected layers. Some of the features utilized that are incipient approaches to convolutional neural networks are ReLU Nonlinearity, Multiple GPUs, and Overlapping Pooling[5].

ResNet50. ResNet-50 is a convolutional neural network that consists of 50 layers. The pre-trained network can relegate images into 1000 object categories, such as a keyboard, mouse, pencil, and many animals. As a result, the network has learned opulent feature representations for a wide range of images. The network has the input size of the image as 224x224[5].

Comparison Result of Different Pretrained Models on the same dataset



This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.10527

B. Mask R-CNN

The Mask R-CNN allows in identifying pixel to pixel delineation for detecting a particular part from a whole image. Mask R-CNN consists of mainly two components 1) BB object detection and 2) Semantic segmentation. In the detection part, Mask R-CNN uses almost similar architecture as Faster R-CNN. But in Mask R-CNN instead of using ROI pooling, it uses ROI alignment to allow the pixel to pixel detection and prevent the information losses as possible. For the Semantic segmentation tasks, it uses fully convolutional Networks. Around the BB objects, FCN creates masks by creating the pixel-wise classification of each region. So in overall Mask R-CNN helps to minimize the total loss of the sections and the following losses at each phase at different levels[8].

C. R-CNN

Thinking of tackling the issue of selecting a sizably voluminous number of regions like in a CNN, Ross Girshick et al. put forward a method where we utilize selective search to find just two-thousand regions from the image and he called them as region proposals. Ergo, now, instead of endeavoring to relegate an astronomically immense number of regions, you will be able to just work with selected regions. These two-thousand region proposals are engendered utilizing the selective search algorithm.

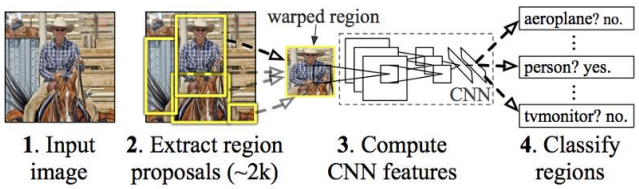


Fig. 2. R-CNN: Regions with CNN features

This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.10527

These 2000 candidate region proposals from the selective search algorithm are warped into a square and victualed into a CNN that engenders a 4096-dimensional feature vector as the output. The Convolutional Neural Network here behaves as a feature extractor and the output dense layer contains the features that are extracted from the image and these features extracted are then alimented into a Support Vector Machine to relegate the presence of the object within that candidate region proposals.

D. Fast R-CNN

This approach is akin to the Region-based Convolutional Neural Network algorithm. But, in lieu of alimenting the region proposals to CNN, we victual the input image to the Convolutional Neural Network to engender a convolutional feature map. From the obtained feature map, we identify the region of proposals and warp them into squares, and by utilizing a Region of Interest pooling layer we reshape them into a fine-tuned size so that it can be victualed into a FC layer. From the Region of Interest feature vector, we utilize a softmax layer to soothsay the category of the region proposals and withal the offset values for the bounding box.

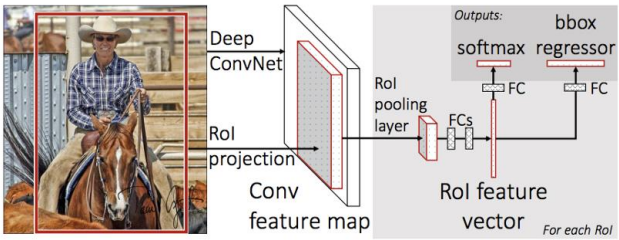


Fig. 3. Working of Fast R-CNN

E. Faster R-CNN

Homogeneous to Fast Region-based Convolutional Neural Network, the image is given as an input to a convolutional network which provides a convolutional feature map. In the first stage, the pre-processed input images are processed utilizing a feature extractor. Then the Region Proposal Network (RPN) will utilize the feature maps as input and outputs a group of rectangular object proposals with their respective scores. The second stage is the Fast R-CNN detector. For each object proposal, the Region of Interest pooling layer will extract a fine-tuned length feature vector from the feature maps. Then each of these feature vectors will be alimented into a sequence of plenarily connected layers to prognosticate the class label and refine the bounding box[7].

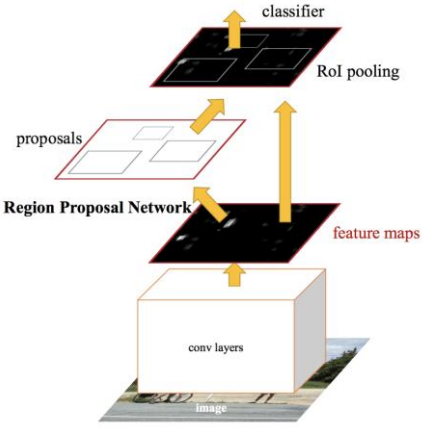


Fig. 4. Working of Faster R-CNN

F. YOLO

All of the discussed object detection algorithms use regions to localize the object that is present in the image. The network does not optically canvass the consummate image. Instead, components of the image have a greater probability of containing the object. YOLO or You Only Look Once is an object detection algorithm that is very different from the region-predicated algorithms visually perceived above. In YOLO a single convolutional network soothsays the bounding boxes as well as the class probabilities for all these boxes[3]. How YOLO works is that we take an image as the input and then split it into an SxS grid, within each of the grid we select m bounding boxes[6]. For each of the bounding box, the network outputs a class probability and offset values. The bounding boxes having the class probability above a particular threshold value is culled and is then used to locate the object within the image.

This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.10527

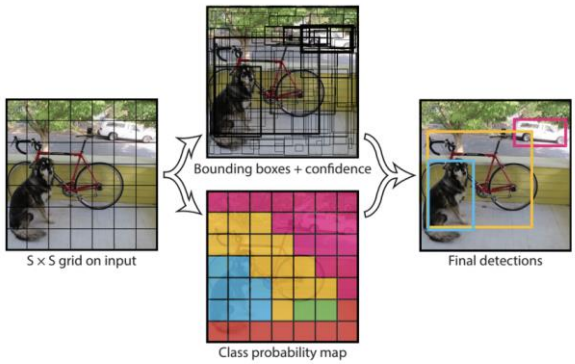
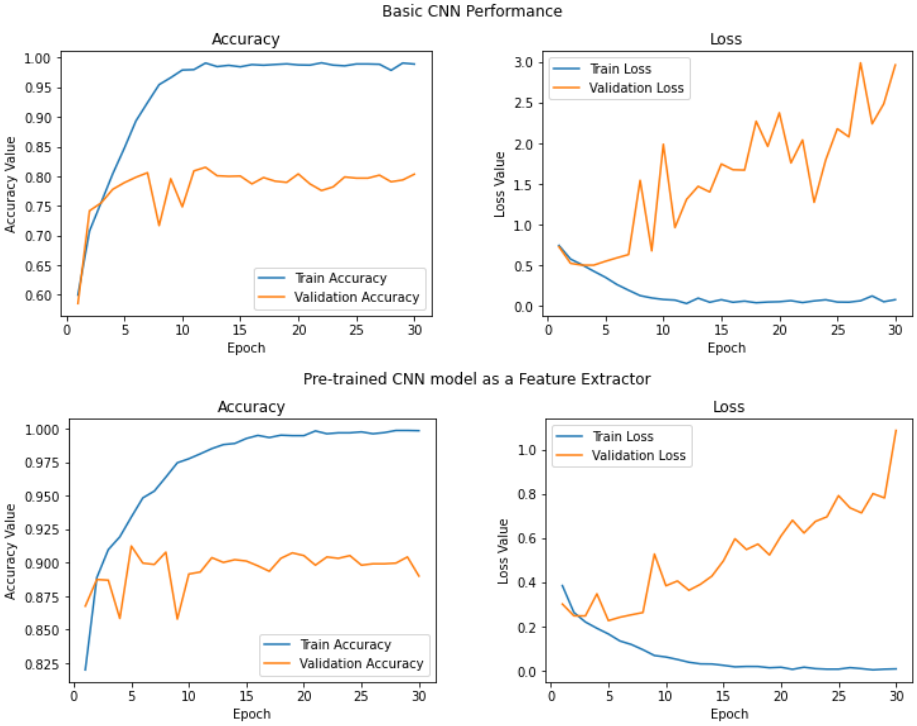


Fig. 5. Working of YOLO Model

Comparison of Transfer learning and CNN on Cat and Dog Classification



III. PREVENTION AGAINST DEEPFAKE VIDEOS

In our proposed AI-driven process of vehicle insurance automation, a customer will use the platform to take video footage of the vehicle's damage. Once uploaded, the system's deep learning model and computer vision distinguish in real-time all the parts of the vehicle, like the roof, window, or bumper and after that spots all the diverse sorts of harm – be it scratch, mark, split, and so on.

With the advent of deepfake videos or AI synthesized videos, it has become really easy to create manipulated videos for misuse and fraudulent claims. So with this automation, it has become really hard to detect whether something is real or fake with untrained eyes like ours. Therefore, a new challenge of detecting deepfakes arises to protect individuals from potential misuses. Also, reports like "Insurance companies lose an estimated US$30 billion a year to fraudulent claims." make it a need of the hour for an automated system with no loopholes.

A. Detection of deepfake videos

For the purpose of deepfake detection, we use a convolutional LSTM architecture which consists of a convolutional neural network (CNN) for frame feature extraction and a recurrent neural network (RNN) for sequence processing and video classification. [1]

This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.10527

A feature vector is obtained for each frame of the test sequence as an output from the CNN. It is then passed to the LSTM structure for further analysis. Finally, the system will predict the probability of the video is a real or manipulated one. The architecture consists of two sub-modules.

CNN for Feature Extraction. For accurately detecting the frame-level features, we are proposing to use the ResNext CNN classifier which is a pre-trained CNN model. We will be fine-tuning the model by adding additional required layers and choosing a suitable learning rate. The feature vectors obtained for each frame are then used as an input to the sequential LSTM. Our model intends to improve on the works done by Abhijit et al. [15] for considering datasets other than facial datasets.

LSTM for Sequence Processing. A sequence of ResNext CNN feature vectors is given as input to the LSTM structure for the classification of the video as deepfake or real. The primary challenge that we need to address is the design of a model to recursively process a Deepfake Video Detection using Neural Networks sequence in a meaningful manner. For this problem, we are proposing the use of a 2048 LSTM unit with a 0.4 chance of dropout, which is capable of achieving our objective. [1][15] LSTM is used to process the frames in a sequential manner so that the temporal analysis of the video can be made, by comparing the frame at 't' second with the frame of 't-n' seconds, where n can be any number of frames before t.

B. Preprocessing

Preprocessing requires splitting the video into frames, followed by the vehicle detection and cropping the frame with the detected vehicle. Frames with no vehicle are ignored as it is unwanted in this scenario.
Previous research conducted by McCloskey et al.[2] has shown that there is a learnable difference between GAN-generated spectral response graphs and camera-generated spectral response graphs. Additionally, they found grayscale histograms to be the most effective in illustrating spectral response variations when comparing different preprocessing methods. A strong difference can be seen when checking this preprocessing approach on a video and it's deepfake counterpart.

C. Architecture

Our model sought to improve on the model created by McCloskey et al. [2] by expanding the input space of our neural network to include a temporal dimension. This was achieved by implementing a 64 neuron LSTM layer into our model. This addition enabled our model to break up each inputted video's 300 grayscale histograms into smaller batches of 10 histograms while maintaining the temporal relationship from the original, larger sequence. The LSTM layer then outputs its results into two more neural network layers that would increment then decrement in size from 128 neurons to 64 neurons, thus ultimately resulting in a final classification for the video. [16]

Fig. 6 shows an overall architecture of the implementation of deepfake detection in our system. After preprocessing and data splitting, the dataset is loaded into our deepfake detection model which consists of the convolutional LSTM structure. The model classifies the input video as real or deepfake.

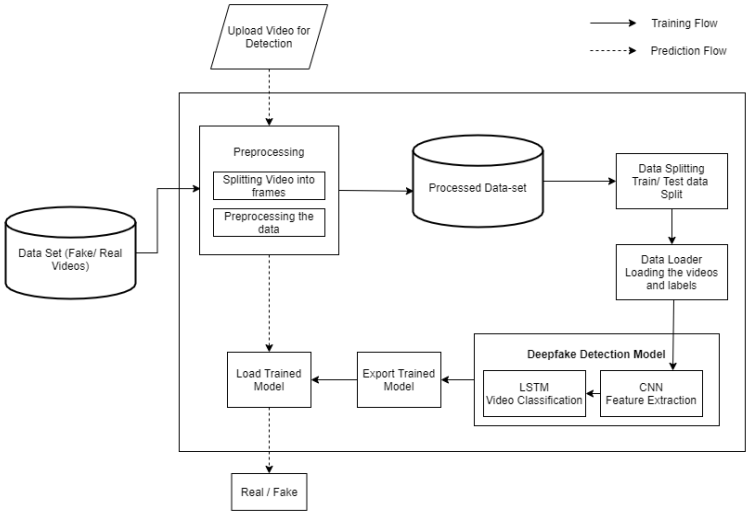


Fig. 6. Deepfake Detection Architecture

This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.10527

IV. PREVENTION AGAINST DEEPFAKE VIDEOS

Images can be manipulated with various image forgery techniques such as image splicing or copy-move forgery. In the image splicing technique, fragments of two or more images are combined to form a single image. Whereas in copy-move forgery a part of the same image is copied and pasted into the same image. Copy move forgeries are more challenging and difficult to detect [11].

CMFD method using SURF and PCET: Image copy-move forgery detection methods are mainly divided into two categories keypoint-based and block-based methods [10].
Wang et al. [11] proposed an efficient CMFD method using features like SURF and PCET and it used the advantages of both the block and keypoints-based methods. Firstly the input image undergoes segmentation and then it is classified into smooth and texture regions.SURF [12] detector and PCET coefficients are applied to obtain the exact keypoints. G2NN algorithm is then used for matching keypoints and the RANSAC algorithm is used for eliminating false matched points. Based on dense points rough rectangular regions are found. Rectangular regions are again divided into overlapping blocks and PCET coefficients are extracted from each block. Later g2NN algorithms are used to find similar features. Finally, we obtain tampered regions by applying mathematical morphology.

A. Preprocessing

Inspired by the CMFD method by Wang et al. [11] using minimum barrier superpixel (MBS) segmentation, the image is divided into non-overlapping irregular image blocks [13]. In order to reduce the scope of looking for similar characteristics, blocks of irregular images are divided into two categories: smooth regions and texture regions. Looking for similar characteristics in smooth regions and texture regions will save more time than in the whole image.

B. Keypoint detection and description

The SURF detector is used to extract key points from smooth and texture areas [11]. This provides key points across the two regions.PCET coefficients are determined and used as descriptors based on each square block associated with the key points.

C. Feature Matching

In the feature matching phase, the improved g2NN algorithm is executed for each group to look for similar features [11]. The g2NN metric is defined with the ratio di/d(i+1) where di is the Euclidean distance with ith nearest neighbor $1 \leq i \leq N$[14]. If the ratio is lower than the threshold two points are matched.

D. False match point elimination

After the feature matching many false points may be present. False match point elimination is achieved via the RANSAC algorithm. RANSAC iteration algorithm based on [15] along with a filtering strate[234]gy combining the label matrix which is obtained by MBS segmentation are being used to find the regions with dense points and hence eliminating the false matched points.

E. Tampered Region Localisation

Minimum and maximum coordinates in x and y directions, for each region with a dense point, is found to obtain a rectangular block. This may not cover the actual tampered region. Inorder to determine the descriptors, PCET coefficients are extracted, and using the g2NN algorithm similar matching points are found. Mathematical morphology close and open operations are being used to eliminate isolated small regions and also fill in holes.

Original image is tampered using photoshop as shown in fig 7.

This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.10527

Fig.7 Copy move tampered image

Fig 8 shows that the copy move forgery was detected clearly. So our CMFD method proves to be accurate and has less run time as described in the method.

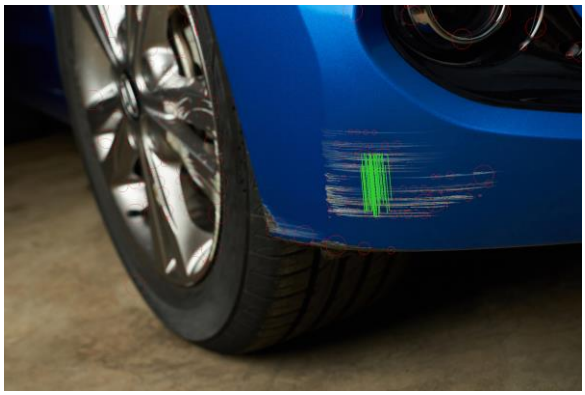


Fig.8 Detected copy move tampered image

V. CONCLUSION

Stating the two most prudent quotes of our time by Andrew Ng, “AI is the new electricity”, and Clive Humby “Data is the new oil”, based on this motivation we amalgamate both data and AI to provide a novel approach for automating the vehicle damage insurance claims. Here out of all the methods discussed for damage classification, we have come to the conclusion that we can implement the same using the combination of transfer learning with cyclic learning rates for training neural networks or by using the YOLO framework. The main challenge is that we have to manually collect versatile dataset from the internet through running web crawler on various search engines like google and bing, and annotate them.Along with the forgery detection model the automatic insurance claim model is proven to be a trust worthy , fast and excellent means of claim processing. From the results and calculations it is expected that this platform would reduce the amount of manual work by about 80 percent and improves accuracy significantly increasing the reliability for the insurance companies.

REFERENCES

[1]. Güera, D., & Delp, E. J. (2018, November). Deepfake video detection using recurrent neural networks. In 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (pp. 1-6). IEEE.

[2]. McCloskey, Scott, and Michael Albright. “Detecting GAN-Generated Imagery Using Color Cues.” arXiv:1812.08247 [Cs], Dec. 2018. arXiv.org, http://arxiv.org/abs/1812.08247.

[3]. Li, P., Shen, B., & Dong, W. (2018). An anti-fraud system for car insurance claim based on visual evidence. arXiv preprint arXiv:1804.11207.

[4]. Patil, K., Kulkarni, M., Sriraman, A., & Karande, S. (2017, December). Deep learning based car damage classification. In 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 50-54). IEEE.

This work is licensed under a Creative Commons Attribution 4.0 International License

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.10527

[5]. Malik, H. S., Dwivedi, M., Omakar, S. N., Samal, S. R., Rathi, A., Monis, E. B., ... & Tiwari, A. (2020). Deep Learning Based Car Damage Classification and Detection (No. 3008). EasyChair.
[6]. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).
[7]. Wang, W., Wu, B., Yang, S., & Wang, Z. (2018, December). Road damage detection and classification with Faster R-CNN. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 5220-5223). IEEE.
[8]. Zhang, Q., Chang, X., & Bian, S. B. (2020). Vehicle-Damage-Detection Segmentation Algorithm Based on Improved Mask RCNN. IEEE Access, 8, 6997-7004.
[9]. S. Teerakanok and T. Uehara, ‘‘Copy-move forgery detection: A state of-the-art technical review and analysis,’’ IEEE Access, vol. 7, pp. 40550–40568, 2019.
[10]. Wang, C., Zhang, Z., Li, Q., & Zhou, X. (2019). An Image Copy-Move Forgery Detection Method Based on SURF and PCET. IEEE Access, 7, 170032-170047.
[11]. P. Mishra, N. Mishra, S. Sharma, and R. Patel, ‘‘Region duplication forgery detection technique based on SURF and HAC,’’ Sci. World J., vol. 2013, Sep. 2013, Art. no. 267691.
[12]. Y. Hu, Y. Li, R. Song, P. Rao, and Y. Wang, ‘‘Minimum barrier superpixel segmentation,’’ Image Vis. Comput., vol. 70, pp. 1–10, Feb. 2018.
[13]. I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, and G. Serra, ‘‘A SIFTbased forensic method for copy-move attack detection and transformation recovery,’’ IEEE Trans. Inf. Forensics Security, vol. 6, no. 3, pp. 1099–1110, Sep. 2011.
[14]. Abhijit Hanumant Jadhav, Abhishek Patange, Hitendra Patil, Jay Patel and Manjushri Mahajan. "Deepfake Video Detection using Neural Networks." International Journal for Scientific Research and Development 8.1 (2020): 1016-1019.
[15]. Pishori, A., Rollins, B., van Houten, N., Chatwani, N., & Uraimov, O. (2020). Detecting Deepfake Videos: An Analysis of Three Techniques. arXiv preprint arXiv:2007.08517.

This work is licensed under a Creative Commons Attribution 4.0 International License