A Very Deep Transfer Learning Model for Vehicle Damage Detection and Localization

Conference Paper · December 2019

DOI: 10.1109/ICM48031.2019.9021687

4 authors, including:

Najmeddine Dhieb
Stevens Institute of Technology

Hakim Ghazzai
King Abdullah University of Science and Technology

Hichem Besbes
École Supérieure des Communications de Tunis

A Very Deep Transfer Learning Model for Vehicle Damage Detection and Localization

Najmeddine Dhieb*§, Hakim Ghazzai*, Hichem Besbes§, and Yehia Massoud*
*School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ, USA.
E-mails: {ndhieb, hghazzai, ymassoud}@stevens.edu
§University of Carthage, Higher School of Communications of Tunis, Tunisia
E-mails: {najmeddine.dhieb, hichem.besbes}@supcom.tn

***Abstract*— Claims leakage is a major problem engendering tremendous losses for insurance companies. Those losses are due to the difference between the amount paid by insurance companies and the exact amount that should be spent, which cost millions of dollars yearly. Experts assert that these losses are caused by inefficient claims processing, frauds, and poor decision-making in the company. With the huge advances in Artificial Intelligence (AI), machine and deep learning algorithms, those technologies have started being used in insurance industry to solve such problems and cope with their negative consequences. In this paper, we propose automated and efficient deep learning-based architectures for vehicle damage detection and localization. The proposed solution combines deep learning, instance segmentation, and transfer learning techniques for features extraction and damage identification. Its objective is to automatically detect damages in vehicles, locate them, classify their severity levels, and visualize them by contouring their exact locations. Numerical results reveal that our transfer learning proposed solution, based on Inception-ResnetV2 pre-trained model followed by a fully connected neural network, achieves higher performances in features extraction and damage detection/localization than another pre-trained model, i.e., VGG16.**

***Index Terms*—Damage Detection, Deep Learning, Insurance, Transfer Learning.**

I. Introduction

Insurance industry is one of the first industry which invested in innovation, high tech, and Artificial Intelligence (AI) [1]. Nowadays, car insurance companies lose a tremendous amount of money due to claims leakage [2], which is defined as the difference between the actual amount paid by the insurer and the exact amount that should be spent in reality. Often, this difference is caused by an inefficient claim processing, poor decision-making, and/or poor customer service. AI technology has shown remarkable improvement in helping making accurate decision in many fields such as robotics, computer vision, and medical science. Many AI techniques are also designed to help in solving a variety of issues in insurance industry such as analyzing and processing data, detecting frauds, minimizing risks, and automating claim process. Such a technology can be also utilized to automate visual inspection and validation in order to help cope with claims leakage.

AI has proved its efficiency in fraud detection for suspected collusion claims as shown in [3]. However, few researchers worked on developing automated visual recognition services in order to create custom solutions for insurance companies to detect and locate vehicle damage. To the best of our knowledge, only the research presented in [4], where a deep learning-based solution was adopted to detect damage in vehicles. Three different approaches has focused on such a topic: training Convolutional Neural Networks (CNNs) from random initialization, convolution auto-encoder based on pre-training model followed by fine-tuning, and transfer learning based on VGG16 pre-trained model [5]. Despite that, the previous study is limited to the damage identification in vehicles without providing extra details. In addition, it cannot evaluate the damage level or accurately locate it, as it is highly sensible to overfitting.

Damage detection and visualization have been studied in other fields. For instance, the study presented in [6] proposed a deep learning pipeline solution for fine-grained classification of building images taken by Unmanned Aerial Vehicles (UAVs) for damage assessment. An integrated deep learning pipeline was suggested to identify structures followed by a fine-grained damage classification for those buildings. Road damage detection and street monitoring solution was developed in [7] where You Only Live Once (YOLO) pre-trained model was adopted to detect various roads damages types as identifiable objects in images.

In this paper, we present a novel framework to detect, locate, and identify damage severity on vehicles using CNN, transfer learning, and Mask R-CNN techniques. Unlike previous studies, our approach not only detects damages but also identifies their severity levels, localizes, and visualizes them on the vehicle's images. In this context, we propose to use the Inception-Resnet pre-trained model [8], as features extractor where we replace the last flatten layer for classification by other neural networks for damage detection and classification. Since object instance segmentation followed by object detection and classification used for roads and buildings damage detection proved their efficiency in those cases. We propose to use Mask R-CNN not only for its performance in object detection but also for its efficiency in instance segmentation. Data augmentation and regularization techniques are employed

This paper is accepted for publication in IEEE International Conference on Microelectronics (ICM19), Cairo, Egypt, dec. 2019.

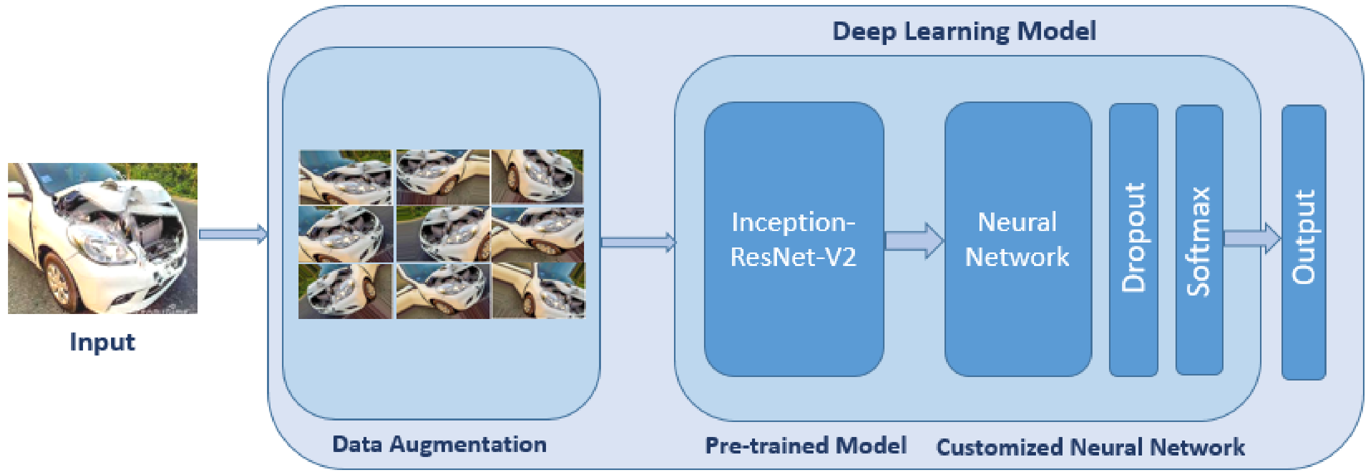

Fig. 1: Model architecture for damages detection and classification.

to reduce the overfitting problems. Then, we provide a comparison between the proposed pre-trained model Inception-ResnetV2 and VGG16 employed in [4]. Results show that Inception-ResnetV2 outperforms the VGG16 due to its large number of hidden layers and its residual connections. For instance, the proposed model exceeds the VGG16 by more than 10% in damage localization.

II. Damage Detection and Classification

In this section, we present our approach to detect, identify, and localize damages in vehicle images. To this end, deep learning, CNN, and transfer learning techniques are used. The workflow in Fig. 1 details our proposed methodology.

- **Dataset:**

Due to the lack of public and accessible datasets for vehicle damage, we manually collect and label images available in the Internet and our local area. We work with tree datasets, the first one contains images for damaged and non-damaged vehicles. The second dataset is composed of three classes to evaluate damage severity: minor (e.g., scratch), moderate (e.g., large dents), and major (e.g., a complete side of the vehicle is damaged). Finally, the third one describes different damage locations. As we are dealing with a small dataset, we use data augmentation to synthetically expand and modify the dataset in order to enhance its performance and relax its tolerance to the overfitting issue during training. We use random rotation, dimension shift, zooming, and flipping transformations to variate the generated data.

- **Transfer Learning:**

Transfer learning is inspired from transfer learned knowledge concept to solve similar problems faster and/or with better performances. It is one of the most effective techniques dedicated for small labeled datasets, where a pre-trained model is used to extract features for the targeted task, while guaranteeing a low overfitting risk. Multiple models pre-trained on Imagenet are available, such as VGG16 and VGG19 [5] as well as Inception [9] are publicly ready for use. In traditional machine learning techniques, the model learns and trains each task from the ground up, whereas transfer learning focuses on extracting features and relevant information from source tasks and applies the acquired knowledge to a target task.

As a result, when the source and the target domains are similar, knowledge transfer may improve the performance of the target tasks. In our case, the source tasks are the classes of the pre-trained model and the target tasks are the damages to be detected, their location, and their severity levels.

- **Model Details and Settings:**

In order to efficiently detect, classify, and localize vehicle damages from images, two major challenges must be addressed: i) the high inter-class similarity ii) and images side and orientation. To build our model, we involve a pre-trained model as a deep feature extractor followed by another neural network to detect and classify the damages. We propose to employ the use of Inception-ResnetV2 [8], a CNN composed of 572 layers and trained on more than one million images from ImageNet, from which we remove the last layer used for object classification. Instead, we add two neural networks, pooling, and softmax layers to accommodate the model to our objectives. A dropout layer is used to improve performance and reduce overfitting problem risk. Since we are using a small dataset, we freeze all the weights of the pre-trained model and only train the last two neural network layers. Regularization parameters are also applied to increase the performance of our model and avoid overfitting.

Since CNN and transfer learning based models are time consuming during the training phase, we use a learning strategy in order to get the best learner parameters in a shorter time. Note that our learning strategy focuses on training the model for k epochs and then evaluate its performance. As long as the validation performance are converging towards the right expected values, we train our model for a longer time, otherwise we adjust the regularization and hyper parameters.

III. Damage Localization and visualization

In this section, we propose to use the Mask R-CNN [10] [11], which is an object detection, classification, and segmentation method, to localize and visualize the damage in vehicle images. We use Resnet-101 feature pyramid network model as a backbone, we initialize our model weights from a pre-trained model based on Microsoft Common Objects in Context (MS-COCO) dataset [12] and train only the network heads to increase our model performance.

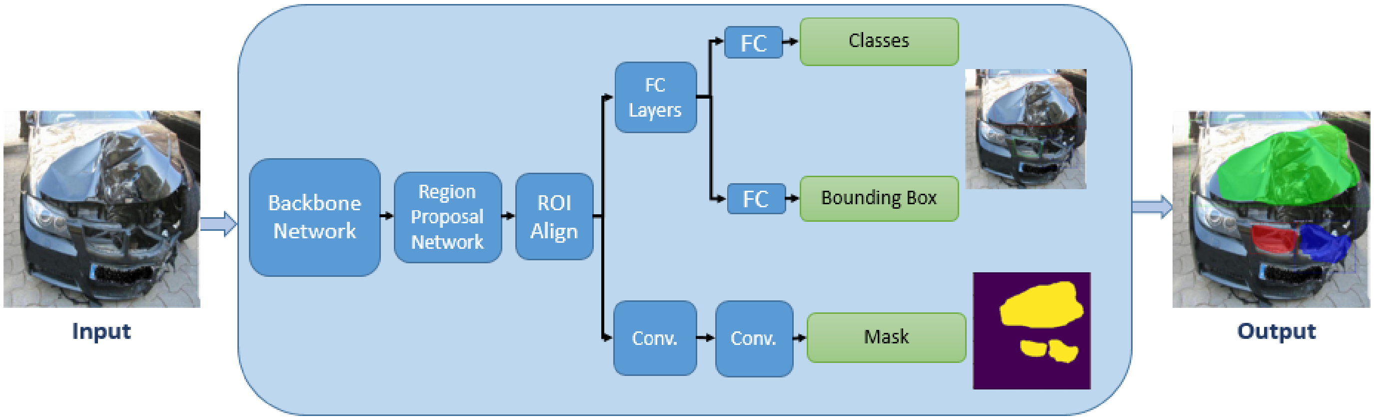

Fig. 2: Mask R-CNN model architecture.

Mask R-CNN is an extension of Faster R-CNN [13], where as a first step a CNN is used to extract accurate features followed by a Region Proposal Network to create Region of Interests (RoIs). Then, a ROI-Align operation allows the construction of instance segment masks. Afterwards, Mask R-CNN introduces another fully CNN to extract the useful and essential segmentation instances in addition to a fully connected neural networks for the classification and boundary box prediction. As a final step, the network heads independently predicts the classes, boundary boxes, and the desired mask.

- **Backbone Network:**

The backbone network in Mask R-CNN model is a CNN used as a features extractor, where low features are obtained from primary layers and deep ones are extracted from later layers. In addition, by passing through the backbone network, images are converted to generate the feature maps. The Feature Pyramid Network (FPN) [14] is a top down pyramid architecture used to extract features from the top pyramid layers and transfer their outputs to the following lower layers, involving lateral connections between the pyramid layers. As a result, each level in the pyramid layers have the access to the higher and lower levels of visual features. In this context, we use Resnet-101-FPN backbone as a feature extractor for the Mask R-CNN to increase its speed and performance.

- **Region Proposal Network:**

The Regional Proposal Network is a fully CNN where the features extracted by the backbone network are used as an input to predict the probability of an anchor being background or foreground. In this context, a sliding window browsing the feature maps generates sets of anchors with different ratios and scales to be used as a bounding box predictor of being a background or foreground object. Since we have an overlap between anchors, we adopt the Non-Maximum Supression (NMS) technique with an Intersection over Union (IoU) threshold set to 0.7 to minimize the redundancy.

- **Region Of Interests Alignment:**

Due to the refinement process of bounding box in RPN, the RoI boxes may have different sizes. In order to get an accurate mask with Mask R-CNN, RoI features should be aligned to maintain the same size as the RoI boxes. Unlike the RoIPool method used in Faster R-CNN, which applies quantification for RoI to discrete the feature map and introduces misalignments between the RoI and extracted features, Kaiming He and al [10] proposed the RoI-Align method where they can avoid the use of quantification and apply the bi-linear interpolation [15] to measure the accurate values of features and hence, the result can be aggregated.

IV. Performance Evaluation

In this section, we evaluate the performances of the proposed AI technique to detect, classify, and visualize damages in vehicles. Two different pre-trained models are investigated in this work, VGG16 [5] and Inception-ResnetV2 [8]. We provide both models the same training and testing data. The models are trained for 100 epochs. Comparisons between both approaches for the damage detection, localization, and severity classification are provided in Table I.

To evaluate the performance of the different transfer learning models, we aim to use three different metrics: Precision, Recall, and F1-score. The higher those metrics are, the best our model is. We observe that Inseption-ResnetV2 pre-trained model outperforms VGG16. Most notably in damage localization and severity classification. The proposed pre-trained model is more efficient in damage localization with the precision of 86.8% in contrast to 75.4%. In addition the VGG16 shows minor performance with only 69% of precision in damage severity classification compared to 80% for the Inception-ResnetV2 model. Overall, the best performances in all challenges are achieved by the Inception-ResnetV2. The notable performance of the proposed model can be validated by the confusion matrices provided in Fig. 3, which summarize the normalized predicted values of each class. The difference between those models can be explained by the fact that the Inception-ResnetV2 has residual connections that allow shortcuts in the model to train very deep neural networks without overfitting problems, which results in better performance.

The Mask R-CNN model performance during training and evaluation processes is evaluated using a multi-task loss function which combines classification, localization, and segmentation masks losses. The classification and bonding-box losses are the same as those defined in [13]. Since there is any competition among classes for masks generation, the segmentation mask loss is defined as the average binary cross-entropy loss, including only the k^{th} mask if the region is associated with the k^{th} ground truth class. The evolution of loss function and mask R-CNN loss function are shown respectively in Fig. 4. The lower the loss function is, the better

TABLE I: Performances for damage severity classification

Metric / Pre-trained Model	Performances for damage severity			Performances for damage localization			Performances for damage detection		
	Precision (%)	**Recall**	**F1-score**	**Precision (%)**	**Recall**	**F1-score**	**Precision (%)**	**Recall**	**F1-score**
VGG16	69.4	0.68	0.67	75.4	0.75	0.75	94.5	0.95	0.94
Inception-ResnetV2	80.2	0.78	0.78	86.8	0.86	0.85	96.8	0.96	0.96

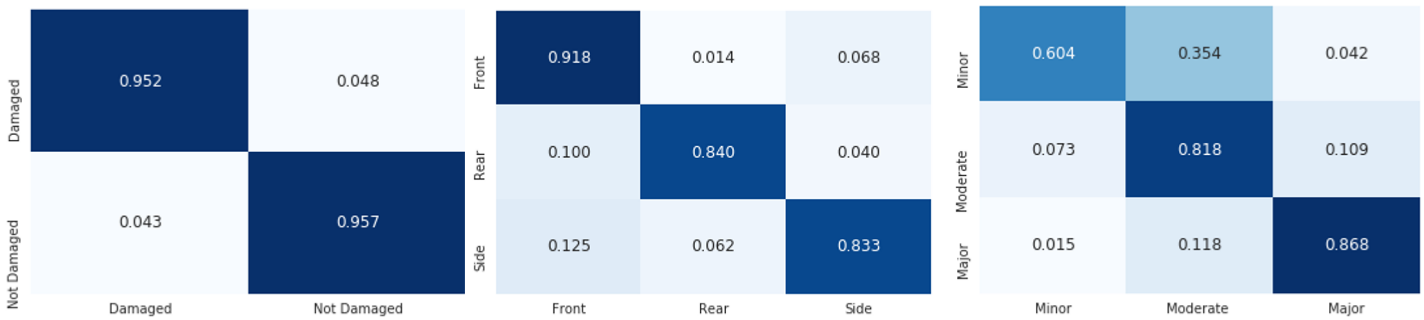

Fig. 3: Confusion matrices for damage detection, localization, and severity classification .

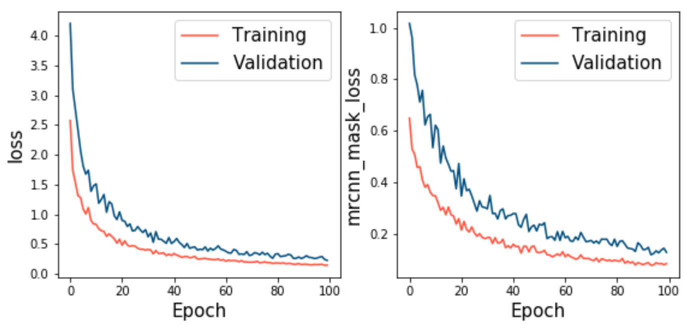

Fig. 4: Loss function evolution.

Fig. 5: Examples of localization and visualization.

a model is. Since the loss functions computing on training and validation datasets are close to zero, we can insure that the proposed model performs well in contouring damages.

Fig. 5 visualizes some examples of the proposed AI-based damage localization. The model can identify damages from different locations and orientations as well as different types and sizes of car damages.

V. Conclusion

In this study, we introduced a novel deep learning architecture to detect, locate, classify, and visualize the damages in vehicle images using transfer learning and instance segmentation techniques. We proposed the use of Inception-Resnet pre-trained model as features extractor followed by fully connected neural networks to identify and classify the damages. We also proposed the use of Mask R-CNN to outline the damage location. The empirical results show that the suggested pre-trained model not only detects damaged vehicles but also identifies its location and severity level. This solution shows an important asset for insurance industry to fight against claims leakage problems.

References

[1] N. Dhieb, H. Ghazzai, H. Besbes, and Y. Massoud, “Extreme gradient boosting machine learning algorithm for safe auto insurance operations,” in *IEEE International Conference on Vehicular Electronics and Safety (ICVES'19)*, Cairo, Egypt, Sept. 2019.

[2] M. Wassel, “Property Casualty: Deterring Claims Leakage in the Digital Age,” Cognizant Insurance Practice, Tech. Rep., 2018.

[3] K. Supraja and S. J. Saritha, “Robust fuzzy rule based technique to detect frauds in vehicle insurance,” in *IEEE Inte. Conf. Ener. Comm. Data Analy. Soft Comp. (ICECDS'17)*, Chennai, India, Aug. 2017.

[4] K. Patil, M. Kulkarni, A. Sriraman, and S. Karande, “Deep learning based car damage classification,” in *IEEE Int. Conf. Machine Learning App. (ICMLA'17)*, Cancun, Mexico, Dec. 2017.

[5] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *arXiv e-prints*, Sept. 2014.

[6] N. Attari, F. Ofli, M. Awad, J. Lucas, and S. Chawla, “Nazr-CNN: Fine-grained classification of UAV imagery for damage assessment,” in *IEEE Int. Conf. Data Sc. Adv. Analy. (DSAA'17)*, Tokyo, Japan, Oct. 2017.

[7] A. Alfarrarjeh, D. Trivedi, S. H. Kim, and C. Shahabi, “A deep learning approach for road damage detection from smartphone images,” in *IEEE Int. Conf. Big Data (Big Data'18)*, Seattle, Washington, USA, Dec. 2018.

[8] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning,” *arXiv e-prints*, Feb. 2016.

[9] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in *IEEE Conf. Comp. Vis. Patt. Recog. (CVPR'16)*, June 2016.

[10] K. He, G. Gkioxari, P. Dollr, and R. Girshick, “Mask R-CNN,” in *IEEE Int. Conf. Comp. Vis. (ICCV'17)*, Venice, Italy, Oct. 2017.

[11] W. Abdulla, “Mask R-CNN for object detection and instance segmentation on keras and tensorflow,” https://github.com/matterport/Mask_RCNN, 2017.

[12] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, “Microsoft COCO: Common Objects in Context,” *arXiv e-prints*, May 2014.

[13] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” *IEEE Trans. Patt. Analy. Mach. Intel.*, vol. 39, no. 6, pp. 1137–1149, June 2017.

[14] T. Lin, P. Dollr, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” in *IEEE Conf. Comp. Vis. Patt. Recog. (CVPR'17)*, July 2017.

[15] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, “Spatial Transformer Networks,” *arXiv e-prints*, Jun 2015.