Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Automobile Insurance Processing using Deep Convolutional Networks

To cite this article: R. Rajkumar *et al* 2021 *J. Phys.: Conf. Ser.* **1783** 012040

Journal of Physics: Conference Series **1783** (2021) 012040 doi:10.1088/1742-6596/1783/1/012040

Automobile Insurance Processing using Deep Convolutional Networks

R.Rajkumar*, Ronhit Neema, N.Chaitanyanathreddy, G.Varun, K.Govinda

School of Computer Science and Engineering, Vellore Institute of Techcnology, Vellore, India

*vitrajkumar@gmail.com

Abstract. We focus on automating the task of automobile insurance processing using Deep Convolution Networks, due to the limited data we find that using transfer learning and using variational Auto encoders to find features works well, we have created four models, classification of car or not, classification of whether car is damaged or not, classification of where the damage has occurred and the severity of the damage respectively. We show different methods that can be used in performing car damage analysis and this paper is not an exhaustive search over the entire domain car damage insurance and claims processing. This paper records the different techniques we have employed in order to analyze the car damage.

1. Introduction

In the automobile insurance sector, there is a lot of money that is wasted due to the leakage of claims [1][2]. Multiple methods have been introduced such as visual inspection in order to reduce this claim leakage however these new methods introduce another problem which is delay in claim processing and a significant amount of time and money is wasted in processing these claims. Few companies and startups have made efforts to mitigate claim processing time [3][4]. In order to automate this task, we employ Convolutional Neural Network (CNN) based methods for classification of different images that are required for automation of the entire process. We divide the problem in to five main classification problems in order to reduce the load of learning on individual Neural Network as we have less data. We have created five models for the following tasks respectively, classification of car, classification of whether car is damaged or not, classification of where the damage has occurred and the classification of severity of the damage respectively. To the best of our knowledge there is no publicly available dataset therefore we have created our own dataset. In the second model we try to classify whether the car is damaged or not and we employ transfer learning techniques for this, the third model we try to classify where the location of the damage is front, rear or to the sides, in the fourth model we try to find out the severity of the damage by classifying into three classes low, mid and high damage respectively.

.

2. Related Works

There has been a lot of advancements in the field of deep learning and the field is growing at a very fast pace, Deep learning has also been consistently giving better results as compared to machine learning techniques in variety of tasks [5][6]. Lot of applications like structural dam- age assessment [7]. Authors propose deep learning-based method for Structural health monitoring to find and characterize the damage in forms of composite material.

Table 1. Table shows the classes and the train and test sizes for First Model

Classes	Train Size	Test Size
Front	419	73
Rear	288	50
Side	272	48

Supervised methods require a lot of data and compute sources. Unsupervised techniques need to be employed when the data is less, techniques such as Variational Auto Encoders[8] have been used to learn complex features from the small data and we can use these features for high level classification tasks. AutoEncoders [9] in general have improved to improve generalization performance of the classifier in case of small number of samples.

Transfer learning techniques have helped reduce a lot of effort of creating new architectures and have helped extract important features from the dataset at hand [10][11]. There are many CNN models trained on ImageNet which are publicly available VGG-16[12], VGG-19 [12], Alexnet [6],

Inception [13], and Resnet [14]. We use transfer learning techniques because of two main reasons,

1. Lack of substantial amount of data.
2. The features that are transferred minimizes the effect of over-fitting in case of labeled set [10]

3. Proposed Method

There is no publicly available dataset for the set of classification problems that we have, and we have created a dataset obtained from google images and have manually labeled each of these images. We have also segregated the seimages based on the image classification task and have created a custom dynamic data augmentation pipeline which will randomly crop and rotate the images all the code is written in PyTorch and is available for reproduction publicly [20] for synthetic data creation in order to improve model performances.

This paper will use the SLR approach to review research on the Naïve Bayes algorithm with the problem of attribute independence assumptions. Systematic Literature Review (SLR) is a process for identifying, assessing, and interpreting all available research with a view to providing answers to specific RQs[15], . In the guide that Kitchenham has made in 2007[15], the literature review will be compiled based on the Systematic Literature Review.

Table 2. Table shows the classes and the train and test sizes for Second Model

Classes	Train Size	Test Size
Damaged	920	
Whole	920	30

Table 3. Table shows the classes and the train and test sizes for Third Model. In this section we have documented the types of models used and the results that we have achieved using these models.

Classes	Train Size	Test Size
Minor	278	48
Modera te	315	55
Severe	386	68

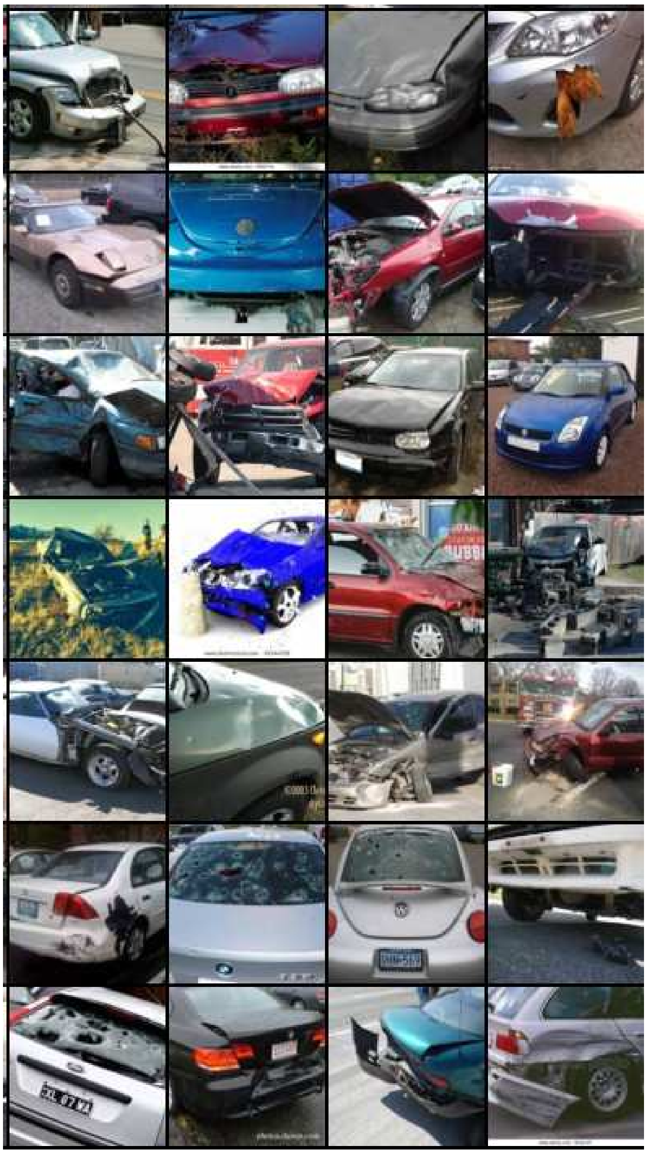

Figure 1. Sample from the dataset

Journal of Physics: Conference Series **1783** (2021) 012040 doi:10.1088/1742-6596/1783/1/012040

The task of the first model is to classify whether the given image is of a car or not, we do not train any model for this and use a pre-existing SOTA model trained on ImageNet in particular we have used the Inception V3 model. Image net consists of 21 classes of cars and by lumping all these categories as cars we have simplified the pipeline of finding whether the given image is of a car or not. If the model classifies the image to be part of any one of these classes then it is classified as a car and it is not classified as a car.

4. Classification of Damage

For this problem we have mentioned the amount of data that we have in Table I, we have also created a custom PyTorch transformation pipeline for data augmentation and improving model's performance. For this challenge we have applied transfer learning on a VGG19 [12] with Batch Normalization with improved performance. Weachieveaf1scoreof0.92, which is quite good for the amount of data that we have. We have also tried other experiments with VGG16 [12] with Batch Normalization but the accuracy was too low to be of use.

5. Classification of Location of Damage

We need to know where the damage exists on the car, therefore it is necessary to localize the damage, there are different ways to achieve this but we decided that its best to first classify the damage instead of something like instance segmentation, we can segment the damage after we have found the place where the car was damaged. The dataset is seen in Table II. We have segregated the dataset into 3 locations namely front, rear and side, we have trained a pertained Resnet152[14] using transfer learning techniques and achieved an accuracy of 0.70. Due to the variation of dataset and lack of dataset we believe this is the best accuracy we can achieve.

6. Classify severity of damage

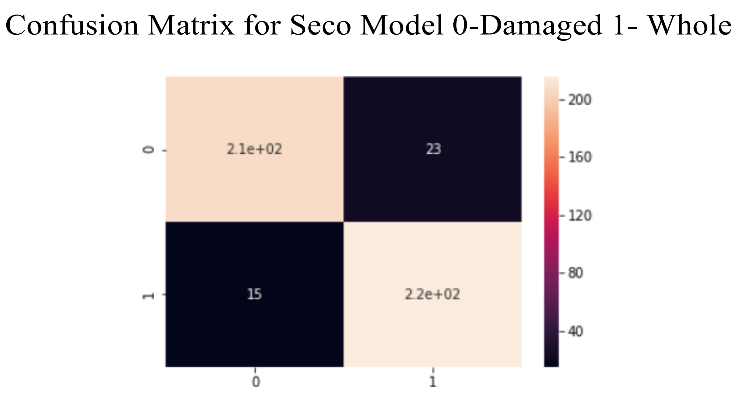

Figure 2. Confusion Matrix for Second Model 0-Damaged 1- Whole

In order to assess damage cost we need to first assess the damage that has been done, we believed that if we can classify the damage to a label it would help in the further analysis of the cost of reimbursement from the car insurance companies. The dataset details are shown in Table III. We have trained a Resnet152 [14] with transfer learning and achieved an accuracy of 0.83.

Journal of Physics: Conference Series **1783** (2021) 012040 doi:10.1088/1742-6596/1783/1/012040

7. Variation Auto Encoders

The amount of variation in the data was huge as a result the VAE was not able to generalize well enough, we couldn't train a deeper VAE because of the lack of GPU resources and data. **Fig 5** shows the images that were reconstructed by the model. FIG. 2. Confusion Matrix for Second Model 0-Damaged 1-Whole

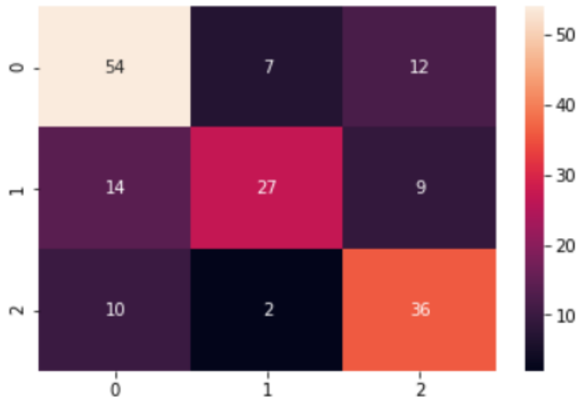

Figure 3. Confusion Matrix For Third Model 0-Front 1-Rear 2-Side

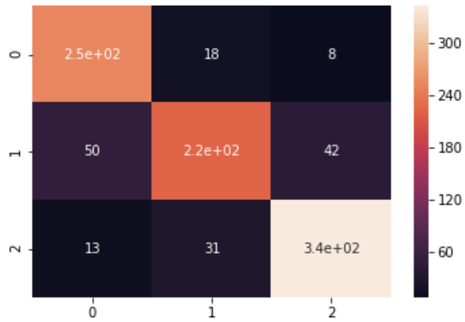

Figure 4. Confusion Matrix for Fourth Model which detects the level of damage 0-minor 1-moderate2-severe.

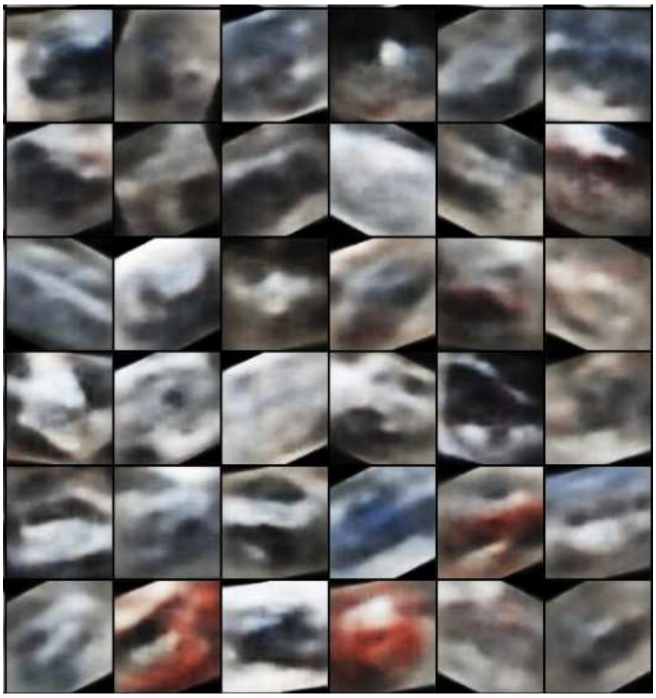

Figure 5. Images Reconstructed using Variational Auto En- coder for 140 epochs on entire dataset

8. Results and Discussion

The contents of the journal are peer-reviewed and archival. The journal publishes scholarly articles of archival value as well as tutorial expositions and critical reviews of classical subjects and topics of current interest.

Authors should consider the following points:

1) Technical papers submitted for publication must advance the state of knowledge and must cite relevant prior work.
2) The length of a submitted paper should be commensurate with the importance, or appropriate to the complexity, of the work. For example, an obvious extension of previously published work might not be appropriate for publication or might be adequately treated in just a few pages.
3) Authors must convince both peer reviewers and the editors of the scientific and technical merit of a paper; the standards of proof are higher when extraordinary or unexpected results are reported.

Because replication is required for scientific progress, papers submitted for publication must provide sufficient information to allow readers to perform similar experiments or calculations and use the reported results. Although not everything need be disclosed, a paper must contain new, useable, and fully described information. For example, a specimen's chemical composition need not be reported if the main purpose of a paper is to introduce a new measurement technique. Authors should expect to be challenged by reviewers if the results are not supported by adequate data and critical details.

9. Conclusion

As we have presented the work in this paper serves to not completely solve the problem of automation of car damage insurance but serves as a foundation of steps one needs to do before any automation of the process is done. We have presented basic key foundation steps that are needed to be done in order to further analyze the image of the car and make meaningful decisions. The models that have been present adhere are highly representative of the dataset and the dataset that has been collected is small and surely not big enough to be representative of wild images that may be found but the core concept behind the training and testing of the semodels can be inherently used for training and testing on custom datasets.

References

[1] JimKremerJennyKillgore.” does your organization need to complete a claims leakage study?”,2015.

[2] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. ”gradient-based learning applied to document recognition”. Proceedings of the IEEE, 86(11):2278–2324, Nov 1998.

[3] Alex Krizhevsky, I Sutskever, and G Hinton. ”imagenet classification with deep convolutional neural networks”. pages 1097–1105, 012012.

[4] Soumalya Sarkar, Kishore Reddy, Michael Giering, and Mark Gurvich. ”deep learning for structural health moni- toring: A damage characterization application”. 102016.

[5] Max Welling Diederik P Kingma. Auto-encoding varia- tional bayes. 52014.

[6] Jonathan Masci, Ueli Meier DanCiresan,andJu¨rgen Schmidhuber. ”stacked convolutional auto-encoders for hierarchical feature extraction”. Pages 52–59, 062011

[7] Maxime Oquab, Leon Bottou, Ivan Laptev, and Jose fSivic. ”learning and transferring mid-level image rep- resentations using convolutional neural networks”InProceedingsofthe2014IEEEConferenceonComputer Vision and Pattern Recognition, CVPR ’14, pages1717– 1724, Washington, DC, USA, 2014. IEEE Computer Society.

[8] Karen Simonyan and Andrew Zisserman. ”very deep convolutional networks for large-scale image recognition”. CoRR, abs/1409.1556,2015.

[9] Govinda.K, Sahaj Singh Maini, “Stock Market Prediction using Data Mining Techniques“, *in the Proceedings of* International Conference on Intelligent Sustainable System, Dec’2017.

[10] K.Govinda, Shruthi Hiremath, “Rainfall Prediction Using Artificial Neural Network”, International Journal of Applied Engineering Research,Vol:9,No:23, pp: 21231-21241.

[11] Rajkumar Rajasekaran, Govinda k, Ashrith Reddy, Uday Sai Reddy, Yashwanth Reddy: *Visual Analysis of Temperature Time Series and Rainfall Using Big Data*. DOI:10.36872/LEPI/V50I3/201023.

[12] Rajasekaran Rajkumar, K.Govinda, Anushka Jindal, Rushil Mehtani, Jolly Masih, Maddineni Charan Sai: *Visualization Effect of Changing Climatic Trends on Natural Calamities:Hurricanes*. DOI:10.36872/LEPI/V50I3/201022.

[13] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. ”rethinking the inception architecture for computer vision”. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 2818–2826,2016.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. deep residual learning for image recognition”. arXiv e-`prints, page arXiv:1512.03385, Dec2015.

[15] B. Kitchenham and S. Charters, “Guidelines for performing Systematic Literature reviews in