SMU Data Science Review

Volume 5 | Number 1 Article 3

2021

Modeling and Application of Neural Networks for Automotive Damage Appraisals

Fred Poon
Southern Methodist University, fpoon@smu.edu

Yang Zhang
Southern Methodist University, yangzhang@mail.smu.edu

Jonathon Roach
Southern Methodist University, jdroach@mail.smu.edu

David Josephs
Southern Methodist University, josephsd@mail.smu.edu

John Santerre
Southern Methodist University, jsanterre@mail.smu.edu

Modeling and Application of Neural Networks for Automotive Damage Appraisals

Fred Poon, Jonathon Roach, Yang Zhang, David Josephs, John Santerre, PhD
Masters of Science in Data Science, Southern Methodist University,
Dallas, TX 75275 USA
PDA Corporation,
Fort Worth, TX 76102 USA
{fpoon, jdroach, yangzhang, josephsd, jsanterre}@mail.smu.edu

Abstract. The automotive damage appraisal process is one of the areas in property and casualty insurance that can benefit from applying deep learning technology and computer vision. It is commercially beneficial to introduce a fast and efficient claim process that can shorten the entire process. Technologies adopted include advanced neural network algorithm and Mask R-CNN to solve tasks such as image classification, object detection, and segmentation in combination with statistical analysis and model construction of the appraisal metadata to approximate final claim cost. With a database of over 3 million records as the data source, a workflow is constructed via a combination of image modeling and Estimate Management Standard (EMS) modeling which can automate the appraisal process with a one-stop approach. It will increase both the efficiency and the accuracy of the process by reducing the labor cost by 50%, as well as reducing the appraisal timeline from an industry standard five days to just hours on average.

1 Introduction

Just a few years ago, the term "insurtech" was barely recognized. Today, insurtech has gained substantial momentum in the US, with a few startups launching mobile applications and machine learning models that can make the insurance claims process more automated. Examples such as Lemonade Insurance, Root Insurance, and Metromile emerged recently, and have gradually transitioned the mindsets of insurance executives to adoption of machine learning and AI. Partnership between technical experts and insurance experts has become the new standard.

One aspect of the insurance workflow still has not quite developed at the same pace. The appraisal process finds itself lagging behind the rest of the industry for several reasons: (1) low accuracy of the existing auto-appraisal tools, (2) the industry's lack of high quality and representative data sets in the US market, and (3) hurdles that exist in insurance regulations and guidelines to link the cost.

Several tools have emerged to address these issues from traditional 3D CAD models to recent innovations in computer vision. Modeling through imagery has proved in many forms to be the method to use today. Technologies involving lasers and 3D modeling have gained little traction due to the price of laser hardware and lack of immediate access to 3D blueprints ([9]). Meanwhile computer vision technologies have been in the industry much longer than lasers, and cameras are inexpensive. Despite this, validation steps are necessary for image classification and pattern recognition results due to concerns of accuracy. The accuracy of such tools varies greatly, and it appears that in some circumstances, the auto-appraisal application is not always reliable.

The appearance of deep learning technologies shows great potential. Image classification using CNN and advanced architectures (e.g., VGG16, YOLO) have been applied to detect car types and vehicle parts that are damaged. From R-CNN [5] to the most recent Mask R-CNN [8], deep learning advanced quickly by doing object detection and image classification, with semantic segmentation tasks. With recent application of Mask R-CNN, the detection and masking accuracy are both greatly increased [21].

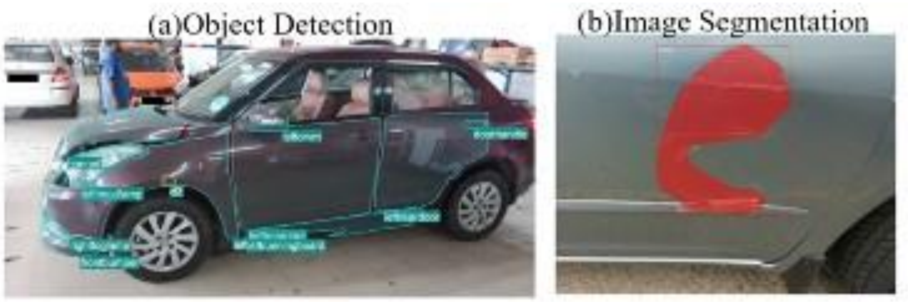

Fig.1. (a) Example image with car parts detection **(b)** Example image with damage area segmentation
(Source: http://ai.stanford.edu/~jkrause/cars/car_dataset.html)

For this research, both CNN and Mask R-CNN are used for image process with adjustments made on the inputs and outputs. Best results are selected with accuracy for image classification, and average precision (AP) scores for mask detection. These measurements come from an integrated approach applied to the automotive damage area with standard photo formats taken after an accident occurs. The system can run through an auto-appraisal process after the photos are uploaded. With this system in place, insurance companies can use this model workflow to process an automobile claim much faster than it did without automation.

In the U.S., there are three major platforms for estimates. An estimate is a line-by-line detailed description of what a vehicle will experience during its time at the body shop. This is how the insurance company and body shop agree to the cost of repair. If there are any disagreements, a new estimate (sometimes called a supplement) is written. An industry standard called Estimate Management Standard, also known as EMS, is released by The Collision Industry Electronic Commerce Association (CIECA) that every estimating platform needs to follow.

EMS, which is a collection of files that when combined, will form an estimate. The estimating platforms are required to read and produce these formatted files. PDA Corporation, an independent appraisal company located in Fort Worth, TX shared EMS data in this study. Their network of over 650 appraisers across the US sees over 400,000 vehicles a year. Our dataset contains more than 3,000,000 records prepared in the EMS format, which will be leveraged to create the cost estimation model.

The future picture of “automotive damage appraisals” shows ample room for technological innovation. The industry standard turn-around time of an appraisal took an average of five days to complete. For insurance companies, it is important to transport the car to the body shop as soon as possible to cut costs. When the company does not have to engage in lengthy conversations with the body shops and wait a considerable amount of time for the estimation to be calculated, it will save tremendous efforts as well as time, labor, and capital. With the process transitioning to an automated procedure, the damage assessment can occur immediately after the pictures are uploaded to the server, and the steps of report generation and claim processing can be streamlined as well. The sooner the repairs are done, the faster the consumer returns to the road. In commercial businesses that involve a fleet of vehicles, getting that damaged vehicle back on the road is vital to those revenue streams.

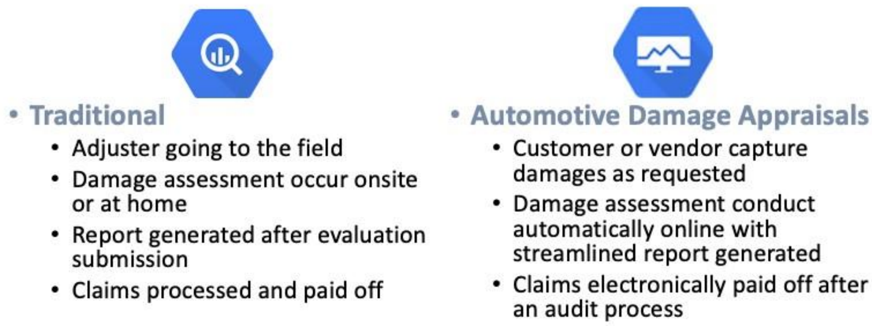

Fig.2. Appraisal workflow comparison

This paper demonstrated that in order to make the “automotive damage appraisals” process work, several components are necessary to successfully deploy into production:

1. Clear and consistent images of the vehicle.
2. A reliable algorithm to detect the types of damages and where they occur.
3. A rich context metadata system that can translate the detailed report containing the parts and severity of the damages into cost.
4. A system that can provide interface to client, auto shop and manual reviewers to reduce uncertainties and allow for quality review.

The structure of the paper is as follows: In section 2, an overview of past tools and technology targeting vehicle damage detection is presented and critically examined. The section 3 is an overview of the available data and methods, including the database from which raw EMS and image data are extracted, and data wrangling steps necessary to prepare the data for both the image modeling and EMS modeling steps. The two methodologies applied in image modeling, CNN and Mask R-CNN, are introduced. Section 4 shows our results of image analysis and accuracy evaluation and benchmarking, as well as the statistical modeling using EMS data together with generated image labels from previous steps. We will conclude with discussion and ethics overview.

2 Literature Review

The expectation is to observe a significant decrease in the amount of time it takes to complete an appraisal after the algorithm is implemented. With an average of five days across the industry, the specific expectation is to see an 80% reduction in time by simplifying the process using the methods described below.

To help pave the way to a faster claim settlement process and increase customer satisfaction, machine learning technology has been used to facilitate the inspection step in recent years. Commercial applications of such technologies [17][18] are beginning to generate tangible results and business revenue. Multiple techniques are investigated, with a target of making the process more accurate and more widely applicable.

Traditional machine learning techniques have also been used to assess the vehicle damages. Jaywardena et al. [9] proposed that 3D CAD models can be used to detect scratch damages from photographs. 3D CAD builds models of ground truth with undamaged vehicles to infer what the vehicles should look like. Image edges are extensively used in 3D CAD model to justify the scratches. Additionally researchers have applied various techniques in solving building and structure damages problems using satellite images in geographical regions [4][10]. The severity of the damage of natural disaster events such as earthquakes can be evaluated from those findings.

The gold standard 3D models can only be provided by OEMs (Original Equipment Manufacturers), such as Honda, Ford, and Toyota. They are essentially hard to come by especially immediately after a new model goes on sale. This presents problems due to the reputation of OEMs for holding onto intellectual property.

Laser technology such as LIDAR has become increasingly popular in recent years. The focus on LIDAR has mainly come with the autopilot features in concept vehicles. There are scanners that could be held and walked around the car to measures distances which can then detect dings and anomalies on the vehicle surface. These aren't ready for everyday consumers and are extremely costly. They are also large in size and unreliable as it brings many uncertainties with its measurements. It remains a possibility especially with consumer grade tools. The fast growth of deep learning technologies opens more opportunities for a more streamlined and automated claim process. In particular, CNN is a powerful algorithm that has wide application in visual object recognition and detection [1]. In 2019, Patil et al [14] proposed the usage of CNN for the purpose of car damage classification. Due to the limited labeled data in their study, they conclude a better result from convolution autoencoder based pre-training transfer learning and ensemble learning, and claim an accuracy of 89.5% using a combined model. They also discovered that only car specific features may not be effective for damage classification and large training set is necessary. In 2020, Malik et al[24] also made use of pre-trained CNN and applied state-of-the-art YOLO (You Only Look Once) object detector to achieve a high accuracy score in a held-out test set.

Several specific research topics are looked into for the car image modeling. One of them is to detect the vehicle object from the background images. In recent years, the area was progressing quickly with many deep neural networks proposed to detect generic objects. Some well-known works include R-CNN [7], Fast RCNN [5], Faster RCNN[16], SSD[19], RFCN[2] and YOLO[15]. In R-CNN [7], Girshick et al. utilized high-capacity convolutional neural networks to help with localizing the objects and did a stepped labeling effort with first scarcely then domain-specific fine-tuning, the methods of regions with CNN features (so-called R-CNN), can give a much simpler but scalable detection algorithm. With another popular approach YOLO [15], Redmon et al framed object detection as a regression problem to separate bounding boxes. At the same time class probabilities are calculated and end-to-end optimization is possible with the detection pipeline. Several improved versions of YOLO are proposed recently as one of the most popular object detection algorithms, this works especially good on recognized multiple different objects from images and videos. Another end-to-end detection approach is to use ResNet [6], Han et al use it to detect vehicle objects in complex traffic conditions and find that residual network can help model to converge using residual modules.

Another important task is image segmentation. Mask R-CNN [8] is the first deep learning network that can do both target detection and segmentation concurrently, which can also be combined and used together with ResNet. Using Mask R-CNN, we can not only accurately detect segments but also possibly label each pixel to different finer features. Mask R-CNN performs better in solving the problems of aligned and multi-scales targets, which could be automobile damaged areas [21].

However, the previous research finds challenges on the transferability of the learning to a more diverse dataset, which is similar to what a European commercial company is facing [2], to implement the identification system to a new geographical region and new market. In our work, we propose a model building process dedicated to the North American market with a high-quality database. Not only do we have standardized data format and images, we also have the necessary components for the metadata model building, which can be used to generate estimates North American currencies. With this unique opportunity, we leverage the recent fast-pace growth network architectures, and the wider application of transfer learning [20]. We construct R-CNN networks [7] the tasks of image classification, object detection and image segmentation and test various pooling layers. We describe our method in the next session and show our benchmark results and share a discussion session with the readers.

2.1 Estimate Management Standard (EMS)

EMS is a collection of 16-17 files that when combined, generates an estimate. Each file in this collection is a subsection of an estimate in text format. The process of wrangling the data involves two steps. First step is to extract the content out of each file. The second step is to transform the data and put it in a tabular format. In this format, the table produces over 200 columns. A problem presented itself when generating the data, as the process created multiple rows in the data to represent one estimate; in other words, for the purposes of our research, multiple rows represent one vehicle. Along with this challenge, a record can be duplicated due to subsequent appraisals or supplements on the same vehicle. This task was a necessary process as supplements could represent additional cost that were not captured originally or reduced cost due to discounts. Having these extra records is not an ideal scenario in the context of predictive modeling.

For early exploratory analysis, the focus is to work with a small subset of the entire data set. The data is filtered on newer generation Honda’s, model year 2015 and newer. This was the only filter, in the vehicle, to retain and allow for a reasonable amount of records for the analysis. Along with this filter, the data was also restricted from the most recent timeline of January 2020 to September 2020. Repetitious records were joined, removed and/or transformed to form unique rows. The most recent supplement or estimate on one particular vehicle was used to avoid multiple estimates on the same vehicle. In its final form, the data set for exploratory analysis consisted of just over 5,000 records.

Our host, Property Damage Appraisers, looks at over 350,000 vehicles a year. With every vehicle, there are at least eight photos. The photos represent the four corners, front, rear, and 2 identifying images of the vehicle. All additional photos would be identified as vehicle damage or supportive photos to the file. The image set goes back nine years. In total, there are over 3,000,000 records for our training and modeling sets. This also means that there are over 3,000,000 images to train an image classification model.

3 Methods

At the study outset, Property Damage Appraisers provided the study team access to an Enterprise Data Warehouse with more than three million claims records, each associated with eight vehicle images, yielding a total of more than 24 million individual data points on which to train models. With this staggering amount of data, preliminary models will need to focus on predicting estimates for a common class of vehicle, such as Honda sedans like the Civic and the Accord. From there, we apply transfer learning techniques to attempt to slowly generalize the algorithms to less common vehicle classes.

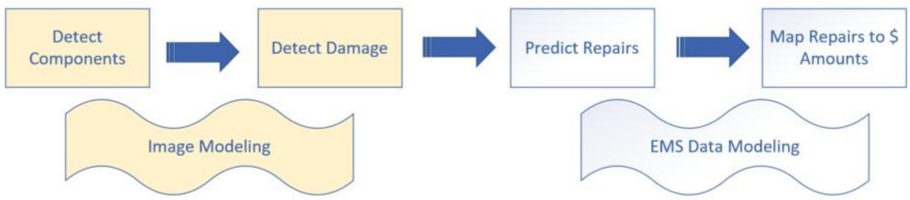

Fig.3. Proposed Model Workflow

During the first phase of the study, the research team's efforts predominantly focused on taking data sets ordinarily used for the purposes of business operations and making them suitable for predictive modeling. This will firstly involve mapping often cryptically named features to more intuitive names using the CIECA Estimate Management Standard. Secondly, the team will be required to develop SQL queries and business rules for wrangling the data into a format that can be used for modeling. At present, the data are in a format that closely mirrors what one would find in a claim estimate report: line items of individual components of the claim. While this format is ideal for readability, it does not lend itself well to any practical form of data analysis. Lastly, the team will turn its attention from the tabular data describing the vehicle and the claim and instead focus on transforming the images of the vehicles into a format that can be meaningfully analyzed by modern programming languages such as R and python.

Due to the fact that the goal of this project is to build a production-ready application, the team faces two fundamental problems related to technical architecture: Firstly, the question of the most suitable method of storing and accessing tabular, non-image data must be addressed. Secondly, the most suitable method for storing and accessing images must also be addressed. The non-image data are stored in a MariaDB relational database and the photos are stored in a remote file system. Each of these storage systems are be hosted on servers at Southern Methodist University's ManeFrame, a center for high-performance computing.

Also related to the goal of the project is the question of which technologies to use train the models and to deploy them to the web. PyTorch is used to train Neural Networks.

3.1 Image Classification

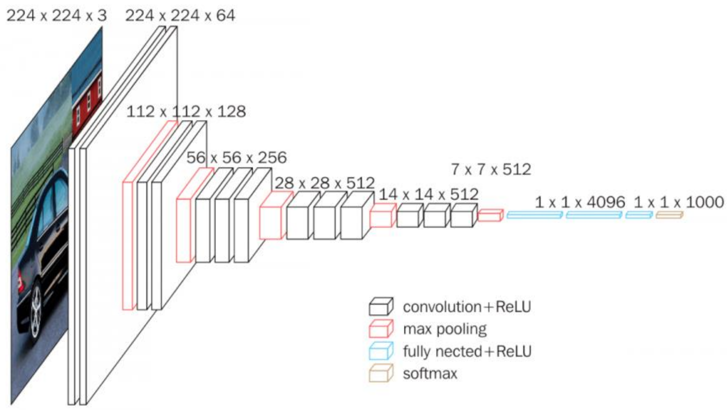

Fig.4. VGG16 Architecture
(Source: http://ai.stanford.edu/~jkrause/cars/car_dataset.html)

VGG16 is a convolutional neural network (CNN) model [22] which is famous for the improvement over Alexnet on image classification tasks. The architecture of VGG16 is composed by convolution layers, 3x3 filters, stride, pooling and padding layers as shown in Figure 4.

The basic components include the follows:

1. Inputs are going through convolution layers of 3x3 filter several times with stride
2. Same padding and max pooling layers of 2x2 filter are used with stride
3. In the end it has 2 Fully connected (FC) layers followed by a softmax as output.

The name VGG16 refers to the 16 layers of the network which have weights, and this network has a pretty large group of parameters (approx. 138 million). The loss function is derived by the misclassified images. Training/validation accuracy and loss are utilized for image classification task to evaluate effectiveness of the algorithm.

3.2 Mask Identification (Mask R-CNN)

3.2.1 Architecture

Mask-RCNN is an extended version of Faster RCNN. There are two main steps it can be divided into: the first step of image scan and general the proposals which are likely to contain objects, the second step is for classification and mask generation. The main step of the algorithm can be described in the five steps below:

Fig. 5. Mask R-CNN Workflow

1. The input images are processed in a pre-trained network. This is the main step for feature extraction and residual blocks are used to conduct the step and store them.
2. Feature Pyramid Network [11] is an extended layer also for the purpose of feature extraction. It is used to represent objects at multiple scales. The concept of it is by using more layers to represent high level features. By doing so it can allow features to have both its lower and higher-level representations. Together with ResNet50, they compose the backbone of our Mask R-CNN implementation.
3. For the next step, the generated feature maps are put into Region Proposal Network (RPN) to generate the bounding box for the Region of Interests (ROIs). The regions defined by the bounding boxes are then put into a classifier to get the categories they belong to. RPN is a lightweight neural network that can scan the image with a consistent sliding-window and find which area contain objects. The detected regions, called "anchors" can be distributed over the whole image area and there could be overlap or covered regions too. RPN can run with the output feature map so it uses the extracted features instead of the raw images, therefore more efficiently in running through. Two outputs from each anchor are expected, one is called anchor class which is the bigger box while another box is the refined box for the object. After running through the RPN process, we pick our top anchors with highest possibilities of containing objects and move to the next step.
4. One the two branches of outputs is the ROI classification process. In the classification step, the position information is further optimized by fit the object more accurately using the boundary boxes. The ROI classifier accepts the ROIs provided from last stage from RPN, it generates both a class of the

object and the bounding box of the object. Notice that in this step, the ROI boxes must have the same sizes. So, a step named ROIAlign [7] is conducted to make sure that the feature map is with the same size, if not, a bilinear interpolation is conducted to reshape the feature map.

5. The feature maps with tighter boxes are also sent to another branch targeting on segmentation masks generating. The mask is a full convolutional network (FCN) [13] that takes the positive regions from the ROI classifier. It generates masks with a low resolution (e.g., 28X28 pixels) to keep the mask branch light. The loss is computed with this light mask. The final masks are sized up to the original ROI bounding box.

3.2.2 Loss Function

The multi-task loss function is defined by a combination of 3 targets:

$$L = L_{class} + L_{box} + L_{mask}$$

The loss from classification and bounding box is the same with Faster-RCNN [5]. The loss function is shown in below:

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{class}}\sum L_{class}(p_i, p_i{}^*) + \lambda\frac{1}{N_{reg}}\sum p_i{}^* L_{reg}(t_i, t_i{}^*)$$

In the above formula the two N indicate the number of classification or regression layers. The two parts are representing the classification or regression loss, respectively. pi is the probability of the object anchor, and p∗i is the ground truth label. ti is the corner representative of the bounding box, and t∗i is the ground truth box label. λ represents the balance coefficient between the two loss functions.

The loss function in the mask branch is defined as the average binary cross entropy loss of all classes. For one class, the loss function mask is only defined on the particular class masks and doesn't affect the loss function based on other masks.

3.2.3 Evaluation Criteria

There are two accuracy measurements we specified to evaluate the performance of Mask R-CNN. We use the average precision (AP) value to evaluate the target detection part, and the mean intersection over union (MIoU) to evaluate the image segmentation part. For our particular task of car-damage detection, we have multiple masks with multiple classes, so mean intersection over union (MIOU) measurements are used together with AP value.

3.3 Transfer Learning

Transfer learning is typically used to solve the data lacking problem. In this study, existing weights of models that already performed high-quality feature extraction are leveraged to save computation time for damage detection. The pretrained model weights utilized in this work are from widely used dataset such as ImageNet [3] and COCO [12]. Although the COCO dataset does not contain a defined car damage class, it contains a large number of images (120K) which contain numerous features that are found common in natural images. These will assist in quickly identifying useful features.

4 Results

In this section we show our modeling results and main findings from the following three tasks:

- Image Classification of car damage severity
- Objective and mask Identification of damaged regions
- Integrated EMS modeling

4.1 Image Classification using VGG16

For image classification tasks we use VGG16 [22] for network architecture. VGG 16 is usually used for image recognition. It's characterized by using only 3x3 convolutional layers stacked on top of each other with increasing depth. We leverage the pre-trained weights by ImageNet and prepared approximately 140 images for image training. The source of the images is a combination of public dataset [23] and our own database.

We group car damages by three levels: Minor, Moderate and Severe. The followed Figure 6 show example car pictures with recognized damage severity. We can see that with proper tagging, the resulted classes are generally align with our intuition. We also found that with features from the pictures, e.g., angle of picture taken, lights, other cars or items in the picture, there are different influences and may interpret the misclassification happened.

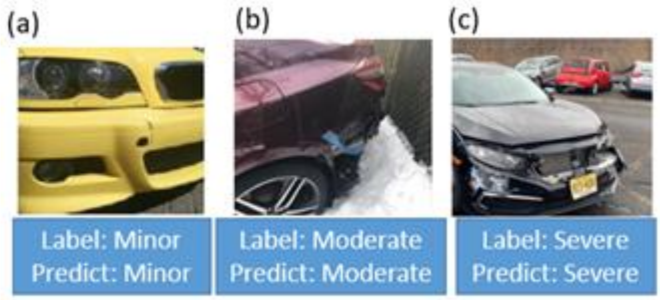

Fig. 6. Example classified car images by damage severity (a) Minor damage (b) Moderate damage (c) Severe damage

With a process to quality check the tagged picture and parameter fine tuning. We used a group of 170 images again from online and local database for initial implementation. The image sub-group is composed from both online pictures and images from our database. The confusion matrix of the best model is displayed in Figure 7.

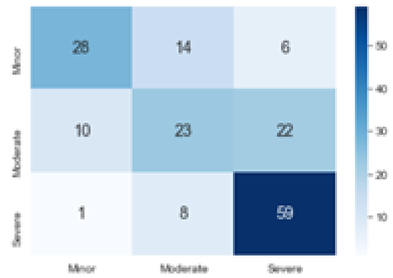

Fig. 7. Confusion Matrix for validation dataset

The over accuracy of the classification is 65.7% for the validation dataset. Specifically, we found that the task of recognized severe damage from the rest are relatively easy (>85%), while differentiating moderate or minor damages is a more challenge task. The accuracy for recognizing moderate and minor damages is <42% and <57%, separately.

4.2 Object Detection and Mask Identification using Mask R-CNN

We conduct experiments on a subset of the whole dataset with 20 cars. To verify that Mask R-CNN can effectively detect type of damages, we set the binary class defined as “damage” or “background” to evaluate the objective detection. Masks are defined as regions with damages.

In order to validate our result we need to execute a process of labeling our dataset. A labeled car image example is shown in Figure 8 marked by VGG Image Annotator (VIA) tool for region and mask definition. The whole datasets are two sub-folders called ”train” and ”val” respectively, to store the labeled images to use for training and testing. The images in each folder corresponded to a file in JSON format.

By incorporating the transfer learning process, some basic features such as color or shape can be borrowed from the pre-trained model. The trained weight files from coco data sets are migrated to dedicated data sets for training. Network parameters are also further fine-tuned during this step. This allows them to achieve accurate results of car damage detection with a relatively small data set, thereby increasing the efficiency of our algorithms and tasks.

Targeting optimization in the car-damage detection task, Mask R-CNN is run on the benchmark data set that is composed of images of 50 cars. By first using the original Mask R-CNN parameters and pre-trained weights from coco data set, the algorithm can be tested by using some sample images from the internet with labels associated with them. The preliminary test results are shown below.

Two example images and preliminary results are shown above in Figure 9. Taking a look at the two images separately, Mask R-CNN performed well in both the object detection and mask segmentation in the first image. The algorithm performed nicely in the second image for object detection of the damages. However, compared to the first example, the mask segmentation task underestimates the area of damages.

Fig. 8. online marked car-damage image

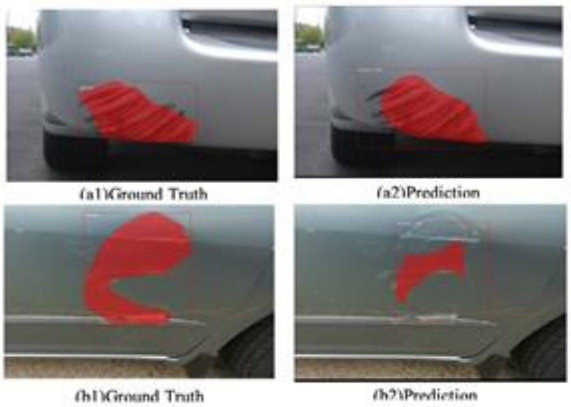

Fig. 9. (a1) Car Damage Ground Truth label with image 1 **(a2)** Car Damage prediction with image 1. **(b1)** Car Damage Ground Truth label with image 2 **(b2)** Car Damage prediction with image 2.
(Source: http://ai.stanford.edu/~jkrause/cars/car_dataset.html)

Although the second car recognized the damages in approximately the right position, the damage area estimated is clearly underestimated. To optimize the model results, further experiences are conducted in two main directions, model weights and model hyper-parameters.

For the initial experiment the weights from coco dataset are used. This Microsoft dataset contains 91 common object categories (including cars) and a total of 328,000 images. It's believed to be a good model to leverage but also considering other model weights that can be used. ImageNet is an image database that is publicly available, which contains 14 million images and are widely used in solving object detection program. This is our second prioritized model to choose for transfer learning experiments.

Fig. 10. (a) Car Damage Ground Truth label. (b) Car Damage prediction with Coco model weights. (c) Car Damage prediction with ImageNet model weights.
(Source: http://ai.stanford.edu/~jkrause/cars/car_dataset.html)

Both experiments above in Figure 10 are using the same hyper-parameters (e.g., Epochs = 30). It can be observed a difference with the predicted car damages using the weights from two dataset and training processes. Relatively, for this particular car a more accurately defined damage area by using ImageNet weights instead of Coco weights. The optimal choices may vary with sample choices. This testing process verify the weights influences to the accuracy of our model predictions.

The second optimization process is by testing out the hyper-parameters, especially the epoch number. After investigating the training and validation error convergence speed, it's found that increasing the epoch number is an effective way to increase the accuracy for both the objective detection and mask determination tasks.

Experiments with the same car by varying the epoch number is displayed in Figure 11 using Coco weights. With a smaller epoch number of 20, the position of the damage can be located but the area detected doesn't not cover the whole area. The accuracy of the detection area is getting higher with increasing epoch number, which appears to be very good in (c) compared with ground truth. The confidence factor of the objective detection task is increasing too from 0.912 at epoch=20 to 0.997 at epoch=60.

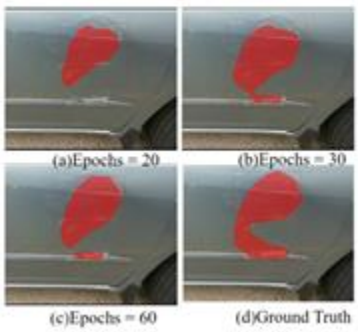

Fig. 11. (a) Car Damage Detection Epoch=20 **(b)** Car Damage Detection Epoch=30 **(c)** Car Damage Detection Epoch=60 **(d)** Car Damage Ground Truth label.
(Source: http://ai.stanford.edu/~jkrause/cars/car_dataset.html)

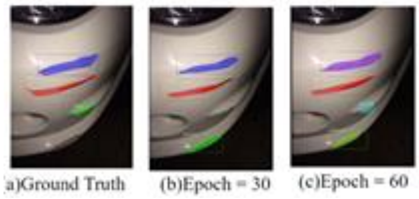

Fig. 12. (a) Car Damage Detection Epoch=30 **(b)** Car Damage Detection Epoch=60 **(c)** Car Damage Ground Truth label.
(Source: http://ai.stanford.edu/~jkrause/cars/car_dataset.html)

Hyper-parameter tuning is not only useful for raising mask determination accuracy, it can also help with multi-target objective detection. The example Figure 12 provides, these are 3 labeled area in the original images. Using our model with epoch number set to 30, one can successfully detect the blue and read areas. However, the model did not find the labeled blue area, instead it finds another damaged area below it. With epoch number increasing to 60, it shows that all 3 original areas are detected with one additional area that is not defined by the original labeled data. It indicates the effectiveness of our method.

4.3 EMS Model Initial Results

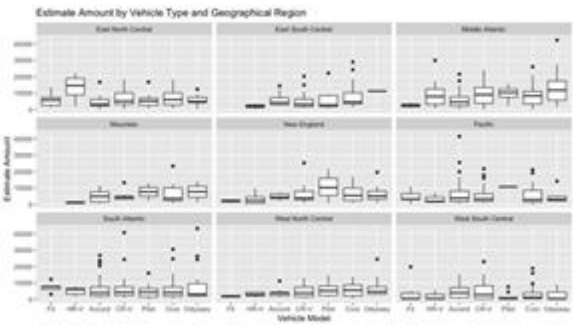

Fig. 13. EMS Data appraisal amount variation Analysis by region

After wrangling the EMS data into a tabular format, a variety of regression modeling strategies, ranging from linear and LASSO regression to Random Forest are used to determine the importance of all the features included in the data set with respect to estimating the cost of a claim. The features initially analyzed include information about the broader context of the vehicle collision, such as the state in which the crash occurred, the state in which the vehicle was appraised, the date of the crash, the vehicle make, vehicle model, and vehicle year, and a variety of other related data elements (shown in Figure 13).

With previous work we can leverage the information we extract from the car images, which is considered an unique source of information compared with other EMS data. We use them combined with EMS data variables as the input features to predict the appraisal amount. The original EMS data with 7256 datapoints and 32 features, using domain knowledge we select 8 of them and one-hot coded the category variables. For demonstration purpose, we narrow the experiment on 183 cars with full set of images (images of 4 corners). With previous established image classification algorithm, we labeled each picture of all the cars with factor levels of 0(Minor), 1(Moderate) and 2(Severe). The 4 labels of LF, LR, RF, RR for each car is included as new features.

Fig. 14. Example car new feature generated from image classification results

With the combined data we conduct random forest algorithm. The first plot of Figure 14 shows the feature importance of the whole dataset without any image labels. The MAE of the resulted appraisal prediction is 2091. The recognized most important features that are dominated are Mileage and Model Year. As we only choose Honda as selected brand to conduct the EMS modeling, branding is not appearing to be a dominate factor.

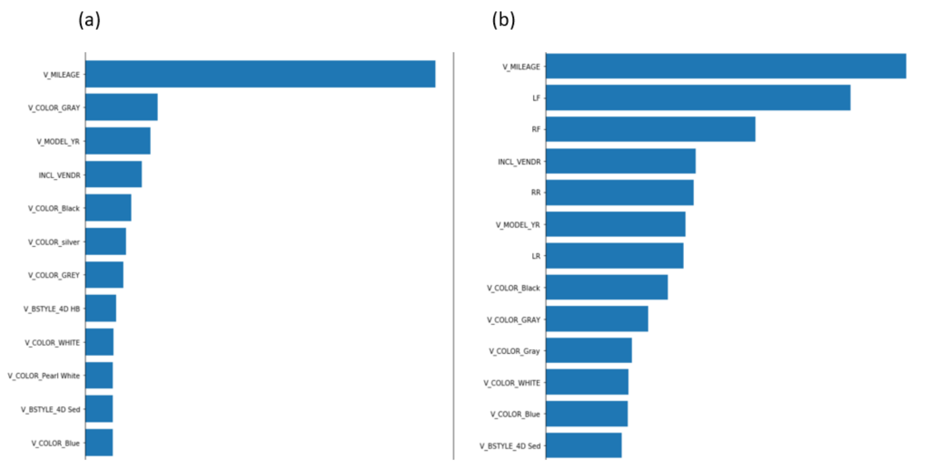

Fig. 15. Feature Importance of (a) original dataset (b) subset dataset with image features added

We conduct the same algorithm to the dataset including the image labels. With 4 new variables we see that the MAE has reduced to 1938. Figure 16 shows the feature importance of the subset using damage severity results from image classification. As displayed, LF is the most important image label recognized of the four, followed by LR, RF and RR. LF is also ranked the 2nd place among all features. It indicates that the model has received additional valuable information. Therefore, it is inspiring for us to systematically expand our experiments to the complete dataset and build a systematic pipeline in the future

5 Discussion

5.1 Breaking the Industry

The industry is only beginning to introduce this technology over the last few years, especially in the automotive insurance industry. It's still unproven in many ways. An estimate is composed of figures and descriptions directly from facilities that provide them. This is a channel the industry hasn't taken advantage of or tapped in to. There are also guidelines and restrictions from the OEMs (Original Equipment Manufacturers), the insurance company, and possibly the State's department of insurance. OEMs can be a decisive factor in the repair process and dictate certain instructions and fits during the repair of the vehicle. Insurance companies further add to the complexity by having their own rules and guidelines on what can be written. They can restrict repair to OEM parts with no aftermarket parts allowed during repair or they can put thresholds and decide that repairing a part is more than sufficient when compared to just replacing a part. Finally, State guidelines can dictate types of appraisals allowed like only to be seen by an appraiser and put in place restrictions where virtual (photo estimating) or algorithmic models are not allowed. This is not the comprehensive list as restrictions and guidelines can change year over year.

Fortunately, laws such as the rights to repair have made its way through state governments. Rights to repair laws bring more control back to the customers and less restrictions from businesses. Though it's still not widely accepted, we have seen more and more state governments heading down this direction.

How do we continue to build reputation and stronger relations with companies that are willing to accept these technologies? With these hurdles in place, our approach is to use a combination of image classification and ranking algorithms to determine an approximation of an incoming appraisal. It's an ensemble type approach with nearest-neighbors' characteristics. As images are submitted, we'd run the algorithm to apply quality assurance to the service and end product. With this method, we can gather more data and build a better reputation. We can provide good example with this process we've observed in this paper. Images are presented to us in real time. Images are tied to an appraisal, or in our case a vehicle. We've already got basic information by knowing what vehicle we're looking at. VIN Decoders can help us identify specific packages your vehicle may have come in.

An interesting observation of the people in the industry has an effect on the future as well. The median workforce age in 2020 for insurance carriers is 44 [26]. As younger generations come into the workforce, we expect this number to drop every few years. But with this, we also expect appraisal experience to drop and a slow rise in machine learning applications to fill that gap.

5.2 Image Modeling

With the optimized model shown leveraging transfer learning weights and hyper parameter tuning, we get good quality model for but the classification and objective and mask identification tasks.

For the data labels as shown in our current Mask-RCNN example results (Figure 13), our model did a good job to detect a fourth area that is not initially labeled. This can be recognized as a success of our model. However, two problems are raised for further exploration and quantification: First, how will the label quality affect our results? Second, should we change our evaluation matrix for situation like this?

The quality of the input images and labels are recognized to be very impactful. As the damage types varies (e.g., dents, scratches, deformations), though we utilized a unified class of “damage” to define it, we have observed that with the types different, the sensitivity levels may change. Another aspect is the quality of the image itself. A few factors that’s recognized that may affect our results including lighting, ambient environment, resolutions and if the picture is squeezed or stretched. Extra efforts may be requested to process those images with questionable quality.

For our application, choosing proper image label to incorporate is a challenging part. The choices are related to the domain knowledge on cars, and also the familiarity with the existence and appropriateness of current available image database applications. In our work, we have a unique source of data, with multiple pictures from the same car. By leveraging the abundance of car information, we evaluate the 4 corners of each vehicle and get a set of image label that can be better representatives of the car conditions. Another information source that we didn’t further discuss is how to utilize the Mask R-CNN results in EMS modeling. We did experiments with ways to incorporate boxes and mask to feature labels, such as mask numbers, mask areas. With its complexity it would need further work, and innovative ways for integration.

6 Ethics

In our research, ethics is mostly an internal conversation. In regard to organizations like the insurance companies, how do protect our client’s information and how do we protect our data? Specifically, to this research, we do not use and personally identifiable information. The key components to our data set are parts, prices, and severity, all information which can be categorical or continuous. We omit key information such as the Vehicle Identification Number (VIN) and license plates from the data. This data doesn’t play much of a role in our data set. And the U.S. is headed down this path regardless as rules and regulations will likely prevent companies from using or sharing this information in the future.

The topic of data privacy here varies in different parts of the world. The European Union have the GDPR, or the General Data Protection Regulation. This is to prevent usage of personal data outside the domain area. It's to give individuals control over their personal data [25]. The GDPR was introduced in 2016 and implemented in 2018. After its introduction, it became a discussion topic at many conferences such as Connected Claims USA and InsureTech Connect. These two conferences in particular brought claims executives together with technology startups. With the opportunity to appear at each of these conferences during these pivotal years, questions were raised by insurance executives. The audience received divided answers from each of the panels. Executives took the approach of "let's sit back and watch". Startup tech company responses were interesting as well. They are going to develop their systems with the GDPR in mind. Technology isn't limited geographically; the technology is effective wherever vehicles are sold. In the following years at these conferences, the tone begins to change to the effect of favoring partnerships between claims industry corporations and technology startups.

Outside personally identifiable information, we also had the ethics of organizational information. There are not many channels to obtain car part information or prices. As mentioned, OEMs and even insurance companies are highly aware and as a result, highly covert when it comes to data. Pitting the startups and corporations is taking the industry in the right direction. As development continues and trends take us in this direction, more statistically drawn models of appraisals will likely be integrated into the workflow.

There's a lot of progress over the years as well. In the homeowner insurance branch of property and casualty insurance, we've been seeing more companies successfully use satellite imagery of roofs and estimating damages. Area coverage is needed when weather events like tornado / hailstorm warnings come through. Satellite imagery takes a before and after photos and compares. Algorithms are running to look for damage and result is produced for the analysts to produce a roof estimate.

Translating this over to automotive appraisals isn't necessarily the same as some people may think. The before and after photos is part of the method of training and testing in the algorithms. But what presents itself during these photos is the environment. In a manufacturing warehouse where a vehicle sits in a controlled environment, we can control the surrounding. We can encase the vehicle in a tent or add more lights to control the lighting of the environment. This allows for consistent photos for every single vehicle coming down the assembly line. Outdoors and parking lots are where most of our dataset came from. Vehicles are reflective with the different types of paints available and this interfered with image processing. Often catching large light 'blotches' as damaged areas can throw off our accuracy.

7 Conclusions

In the paper, we established a pipeline for automatic appraisals that can potentially be scaled to a larger dataset. In this integrated workflow, we focus on processing the database of car images into labels that can represent the car damages from multiple levels and converting them to labels that can be easily digested together with other EMS features for appraisal value prediction.

Despite the challenges we confronted with car images of varies quality, we built an image multi-classification model with CNN to classify car damage severity. The results is most accurate on separating severe damage from the rest (~87%) with sample data. We also demonstrated that using advanced algorithms such as Mask R-CNN, we could extract image information with objective detection and mask segmentation tasks embedded to reflect where and how many damages occurs. We displayed the influences on the final results with transfer learning weights selection and parameter fine tuning and pointed out that with further work there's improve space to extract finer car-related information. A final appraisal price estimation modeling executed with image labeling included has an MAE value reduced to a much lower value, which reflect the potentials that the image labels can bring. This bring both technique and business opportunities to the future roads of automotive appraisals application.

Our dataset has a good representation in the U.S. If deployed, this can fit any business organization in North America. Due to the repair and regulation similarities between U.S. and Canada, we are confident our models can be applied to both countries. We have a good representation of vehicles available as well. Concepts and algorithms can be applied outside these two countries as well. Though accuracy may differ due to the possibility of different state/provincial regulations. The platforms that produce our estimate do not necessarily have to follow the CEICA standard outside of the U.S. which can prevent us from acquiring the data.

The approach we suggest allows for improvements and tuning within the algorithms we've put forth. Doing a side-by-side analysis of estimates helps build confidence with insurance carriers. And we'll see the shift from adjusters and appraisers to more analysts. This will be the result of better systems and resources as well as a new workforce. As generations of younger individuals join the workforce, we expect new machine learning applications to fill the voids left behind.

The future adoption of such an automatic process can change the flow of how an appraisal process will be done. Instead of relying on analysts to produce an estimate in >48 hours, an automatic flow could be implemented with a controllable uncertainty to simplify the process with a reported appraisal price in hours, especially for the cars with serious damages. Customers can get the results back less painful. The use of deep learning techniques in this area will continue to prove its great value in the future.

Acknowledgments. We would like to thank PDA Corporation for granting us the data for this work. We also want to thank Dr. Jaquelyn Cheun, Dr. Jake Drew for their suggestions and valuable discussions.

References

1. A. Krizhevsky, I. Sutskever, G.H.: Imagenet classification with deep convolutional neural networks, pp. 1097–1105. F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc. (2012)

2. Dai, J.: R-fcn: Object detection via region-based fully convolutional networks. NIPS (2016)
3. Deng, J., Dong, W., Socher, R., Li, L., Kai Li, Li Fei-Fei: Imagenet: A large-scale hierarchical image database pp. 248–255 (2009). https://doi.org/10.1109/CVPR.2009.5206848
4. F. Samadzadegan, H.R.: Automatic detection and classification of damaged buildings, using high resolution satellite imagery and vector data, vol. 37, pp. 415–420. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (2008)
5. Girshick, R.B.: Fast r-cnn. ICCV pp. 1440–1448
6. Han, G.: A method based on multi-convolution layers joint and generative adversarial networks for vehicle detection. Proc. KSII Trans. Internet Inf. Syst. (2019)
7. He, K.: Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition pp. 770–778 (2016)
8. He, K.: Mask RCNN. IEEE Int. Conf. Comput. Vis. ICCV, 2980–2988 (2017)
9. Jayawardena, S.: Image based automatic vehicle damage detection. Ph.D. thesis, Australian National University (2013)
10. K. Kouchi, F.Y.: Damage detection based on object-based segmentation and classification from high-resolution satellite images for the 2003 boumerdes algeria earthquake (2005)
11. Lin, T.Y., Dolla´r, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection (2017)
12. Lin, T.Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P.,Ramanan, D., Zitnick, C.L., Dolla´r, P.: Microsoft coco: Common objects in context (2015)
13. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantics segmentation (2015)
14. Patil, K.: Deep learning-based car damage classification. IEEE International Conference on Machine Learning and Applications (ICMLA) pp. 50–54 (2017)
15. Redmon, J.: You only look once: Unified, real-time object detection. CVPR (2016)
16. Ren, S.: Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence pp. 1137–1149 (2017)
17. Smart, K.: Insurance visual claims fight fraud and reduce costs. Sightcall (February 2016), https://sightcall.com/insurance-visual-claims-fight-fraud-and-reduce-costs/
18. Tractable: https://tractable.io
19. Y. Li, H.H.: Research on a surface defect detection algorithm based on mobilenetssd (2018)
20. Yosinski, J.: How transferable are features in deep neural networks? Advances in neural information processing systems pp. 3320–3328 (2014)

21. Zhang, Q., Chang, X., Bian, S.B.: Vehicle-damage-detection segmentation algorithm based on improved Mask RCNN. IEEE Access 8, 6997–7004 (2020). https://doi.org/10.1109/ACCESS.2020.2964055
22. Simonya, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, arXiv preprint arXiv:1409.1556, [2014].
23. http://ai.stanford.edu/~jkrause/cars/car_dataset.html, 3D Object Representation for Fine-Grained Categorization by Jonathan Krause, Michael Stark, Jia Deng, Li Fei-Fei
24. Malik, H., Dwivedi, M., Omakar, S., Samal, S., Rathi, A., Monis, E., Khanna, B., Tiwari, A., Deep Learning Based Car Damage Classification and Detection, 2020, Easy Chair Preprint no. 3008
25. https://ec.europa.eu/info/law/law-topic/data-protection_en, General Data Protection Regulation
26. https://www.bls.gov/cps/cpsaat18b.htm, Labor Force Statistics from the Current Population Survey