Car Damage Detection: Fine Tune a Vision Transformer for Image Classification

Awais Choudhry, Hsien-Pang Hsieh, Yigitcan Yildiz

Introduction and Background

Cars serve as a backbone for mobility in most societies. Over the last years, the number of actively registered cars in Germany has steadily increased, amounting to around 49 million cars in 2024[1]. Car insurance is mandatory for any cars actively used on public roads, and the premiums collected by insurance companies from car insurances has also risen over the years in Germany. Nevertheless, history has shown that the overall payout from insurance firms for car damage claims is consistently very close to the premiums received, for some years even larger than the premium received[2]. Car damage detection and estimation is crucial for insurance industries, and often involves appraisals by third parties. Such appraisals are also paid out by insurance firms as part of the claim processes. Human assessments of damages can vary depending on expertise and ulterior motives, such as to maximize the payout from the insurance, leading to inconsistencies in claim outcomes. Furthermore, manual evaluations can delay claim settlements, inconveniencing customers and increasing the administrative overhead for insurers.

Given the previously highlighted inefficiencies, we identify the potential for a model to be used for car damage detection. Inspired by a team member's experience at AXA, this project leverages the CarDD dataset to explore automated car damage detection. Vision Transformers (ViTs) are well-suited for this task due to their ability to learn both local features and progressively capture global shapes as the layers deepen. This makes them particularly effective for identifying subtle damage patterns than the Large Convolutional Neural Network. This project aims to fine-tune a ViT to initially classify damage categories in the CarDD dataset.

Methodology for Classification

Dataset Analysis

- **Dataset Overview:** The CarDD dataset (https://cardd-ustc.github.io/) contains labeled images of vehicles with various types of damages, categorized into classes such as dents, scratches and cracks.
- **Data Preprocessing:** The images will be resized and augmented to increase model robustness. Data augmentation techniques such as random cropping, rotation, and brightness adjustments will simulate real-world variability.

---

1 https://www.umweltbundesamt.de/daten/verkehr/verkehrsinfrastruktur-fahrzeugbestand#entwicklung-des-kraftfahrzeugbestands

2 https://www.gdv.de/gdv/statistik/statistiken-zur-deutschen-versicherungswirtschaft-uebersicht/schaden-und-unfallversicherung/geschaeftsentwicklung-in-der-kraftfahrzeug-haftpflichtversicherung-137954

Model Design

- **Model Architecture:** A pre-trained Vision Transformer (ViT) model available on Hugging Face's model hub(https://huggingface.co/docs/transformers/en/model_doc/vit) will be fine-tuned on the CarDD dataset.
- **Loss Function:** A categorical cross-entropy loss function will be used to train the model.

Training and Evaluation

- **Training Setup:** The model will be trained using PyTorch or TensorFlow. Hyperparameter tuning will be conducted to optimize learning rate, batch size, and other parameters.
- **Evaluation Metrics:** Metrics such as accuracy, precision, recall, and F1-score will be used to measure performance. A confusion matrix will provide insights into class-specific performance.
- **Comparative Study** The ViT model's performance will be compared to baseline CNN architectures (e.g., ResNet, EfficientNet) to establish its relative efficacy.

Expected Outcomes and Possible Expansion

The fine-tuned Vision Transformer (ViT) model is expected to outperform baseline CNN architectures in classifying car damages, such as dents, scratches, and cracks, thanks to its ability to capture both local and global features. This capability makes it particularly effective in identifying subtle and complex damage patterns. The confusion matrix and class-specific metrics will help us find areas where the model excels or struggles, guiding future improvements. Depending on available resources and time constraints, the project might be extended to include object detection and instance segmentation, providing detailed localization of damages. Additionally, integrating factors like car age and brand in future research could enhance predictive capabilities, such as estimating damage losses. These advancements have the potential to revolutionize car damage detection, improving the accuracy and efficiency of insurance claim processes while laying the groundwork for innovative applications in real-world scenarios.