DETR versus R-CNN Models:
A Comparative Analysis of Performance in Vehicle Component and Damage Detection

Author: Kai Hennig[1,2] Supervisor: Prof. Dr. Elke Hergenröther[1] Co-Supervisor: Prof. Dr. Andreas Weinmann[2]

[1]Departments of Mathematics and Science & [2]Computer Science of Darmstadt University of Applied Sciences

h_da
HOCHSCHULE DARMSTADT
UNIVERSITY OF APPLIED SCIENCES

Motivation & Research Goals

- **Insurance Context:** Essential for accurate detection of vehicle damages in images for insurance claims processing.
- **Advancement in Object Detection:** Transformer-based architectures like DETR [1] introduce global attention mechanisms, outperforming traditional frameworks such as Faster R-CNN [2] in benchmarks like COCO [3].
- **Research Aim:** Assess if results from large, generic datasets translate to smaller, specific datasets in real-world scenarios, focusing on vehicle component and damage detection.

Data

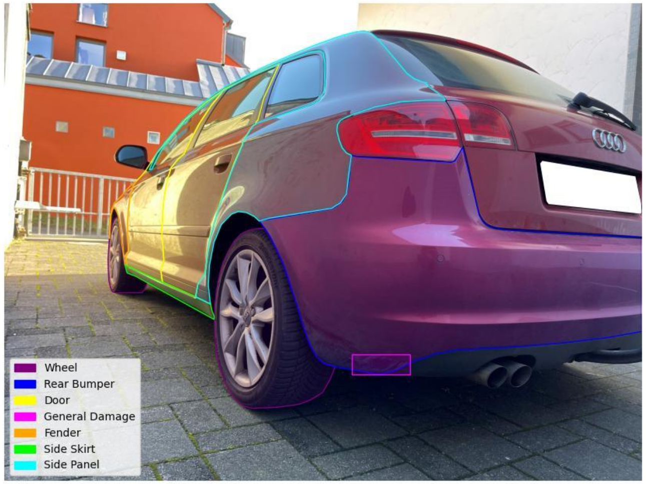
- **Dataset Source:** Custom comprising 7,330 images of damaged vehicles, divided into training, validation, and testing subsets.
- **Image Details:** The dataset includes a variety of image resolutions ranging from 300 x 225 to 9,248 x 6,936, with a median resolution of 1280 x 960. It encompasses both close-up photographs and wide-view images, providing diverse perspectives.
- **Annotation Details:** The dataset is professionally labeled with segmentation for vehicle components and bounding boxes for damages (see image). It includes approximately 100,000 labeled instances. The dataset is divided into Extracted Side Components and General Damage subsets for focused analysis.

Related Work

- **DETR Development:** A transformer-based detection architecture by Carion et al.
- **Benchmark Comparison:** DETR vs. Faster R-CNN on COCO; DETR excels with large objects but underperforms with smaller ones, where Faster R-CNN shows strength; overall, DETR has better performance.
- **Existing Research:** Focused on custom datasets; the recent CarDD [4] dataset provides a base for future studies in vehicle component and damage detection.

Concept and Methodology

- **Approach:** Comparative analysis using Mask R-CNN (trained on Side Components with segmentation), Faster R-CNN (trained on both datasets with bounding boxes), and similarly trained DETR (similar to Faster R-CNN).
- **Implementation Details**: Utilization of pretrained COCO weights, hyperparameter tuning, and no data augmentation.



Comparative Performance Analysis

- **Evaluation Protocol:** COCO detection evaluation metrics**.**
- **Research Findings**: DETR shows superior overall AP, particularly with larger objects. Faster R-CNN is more effective in detecting smaller objects and at lower thresholds (see table)**.**
- **Dataset-Specific Results:** High performance on the Side Components Dataset, with a need for improvement in the General Damage dataset, especially for small objects**.**
- **Comparison to Carion et al.:** DETR's and Faster R-CNN's performance characteristics are consistent; DETR's overall lead narrows**.**

Conclusion

- **Benchmark Translation:** Performance traits of DETR and Faster R-CNN from larger datasets largely reflect in specific datasets.
- **Performance Variation:** Differences across datasets underscore the necessity of tailored architecture testing and selection.
- **Overall Outcome:** Excellent results in Vehicle Side Components detection; substantial scope for enhancement in General Damage detection.

	Side Components		General Damage	
Metric	Faster R-CNN	DETR	Faster R-CNN	DETR
$AP_{0.50:0.95}$	0.631	**0.641**	0.123	**0.131**
$AP_{.50}$	**0.840**	0.830	**0.304**	0.294
$AP_{.75}$	**0.712**	0.694	0.075	**0.107**
AP_S	**0.110**	0.081	**0.005**	0.004
AP_M	**0.371**	0.298	**0.050**	0.034
AP_L	0.681	**0.711**	0.150	**0.165**
AR_1	0.620	**0.626**	0.159	**0.174**
AR_{10}	0.748	**0.776**	0.310	**0.360**
AR_{100}	0.748	**0.789**	0.344	**0.505**
AR_S	**0.245**	0.195	**0.075**	0.052
AR_M	**0.540**	0.472	0.219	**0.237**
AR_L	0.793	**0.852**	0.395	**0.601**

Literature

[1] Carion, Nicolas, et al. "End-to-end object detection with transformers." *European conference on computer vision*. Cham: Springer International Publishing, 2020.
[2] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." *Advances in neural information processing systems* 28 (2015).
[3] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. Springer International Publishing, 2014.
[4] Wang, Xinkuang, Wenjing Li, and Zhongcheng Wu. "CarDD: A New Dataset for Vision-Based Car Damage Detection." *IEEE Transactions on Intelligent Transportation Systems* (2023).