Revolutionizing Automotive Analysis Car Part Segmentation With Deep Learning

Sanem Leblebici (s222448), Michał Lehwark (s222999), Ari Menachem (s163956), Elli Georgiou (s223408)

Technical University of Denmark,

Introduction

Our project explores car part segmentation using two neural network models, Unet and Unet+ResNet, to determine the most effective approach. We improved a Deloitte provided dataset of approximately 3,170 photos using approaches such as data reduction and augmentation to achieve approximately 90% accuracy. This comparison analysis not only increased model robustness but also advanced AI-driven segmentation in the automotive industry.

Dataset

The dataset provided by Deloitte included about 3000 labelled synthetic images and 168 labelled real images. In order to train an effective segmentation model, the dataset at hand needed several modifications. These included reduction of the density of synthetic data, correction of segmentation mask labels and data augmentation. The labels were in numpy array format and the data came with both clean versions and the segmented versions of the images. See these in Figure 2

1. **Repetition in dataset and data reduction:** The synthetic data provided was the majority of the given data and was highly repetitive. This was reduced to a few good scenes.
2. **Mask Correction:** Several of the numpy arrays given as segmentation masks were noisy and incorrectly labelled - these did not match the images provided plotting the segmentations. These were corrected with a pixel by pixel replacement.
3. **Data augmentation:** Data augmentation was used to boost the variety and quantity of the data.

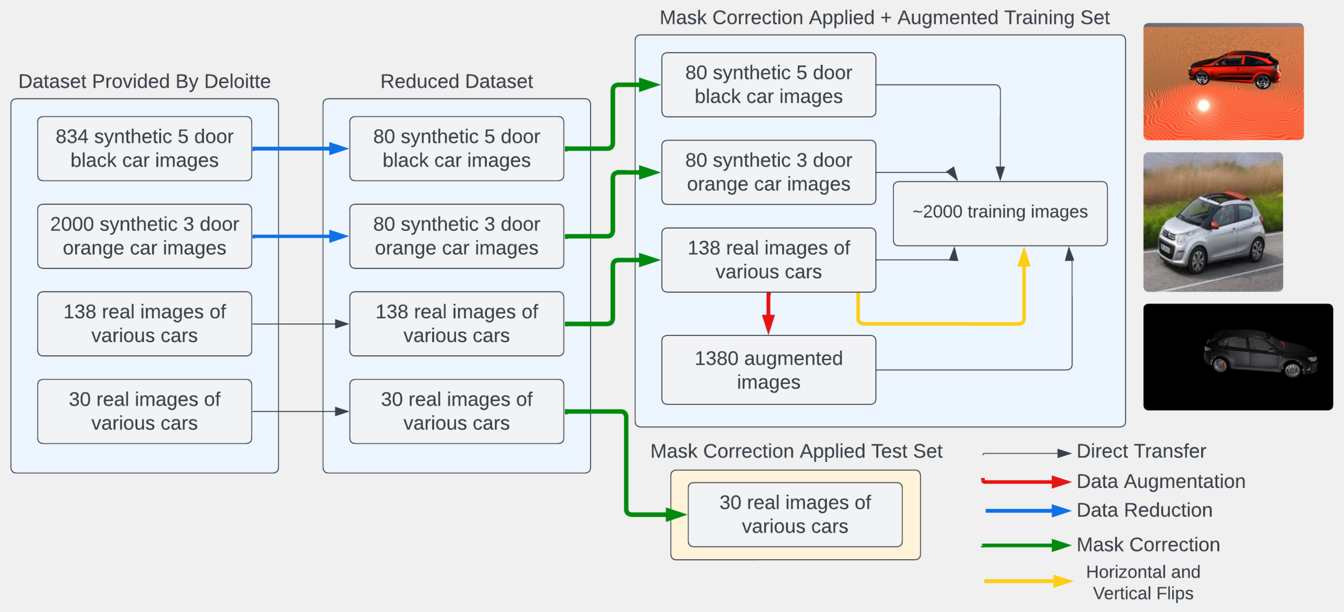
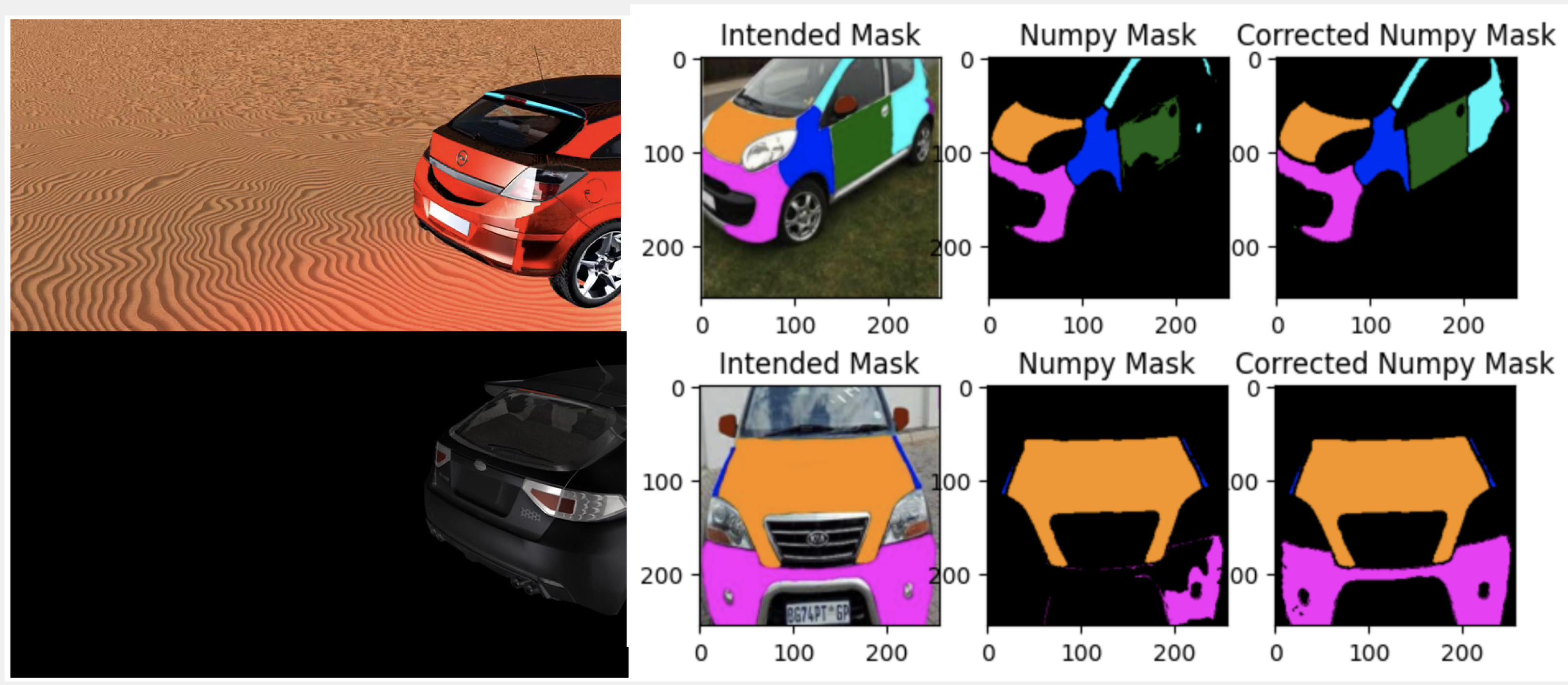


Figure 1. Data pipeline



Examples of excluded images

Examples of corrected masks

Figure 2. Examples of data issues corrected

Augmentation

To increase the size of the dataset, we used an elaborate augmentation technique. Each of the 138 original real car photos were augmented ten times, resulting in a total of 1380 augmented training images.

With this strategy our deep learning models were exposed to and trained on a diversified set of data, considerably improving their robustness and capacity to classify car parts accurately in a variety of real world situations.

Data array	Image noise factor	Color jitter boundary	Occlusion probability	Gaussian blur kernel size
0	0.1-0.2	0.2	0.3	3
1	0.1-0.11	0.2	0.3	3
2	0.5-0.6	0.2	0.3	3
3	0.1-0.2	0.05	0.3	3
4	0.1-0.2	0.5	0.3	3

Table 1. Augmentations added for the specific data arrays

Augmentation	Description
Image noise	Pixel noise added randomly to all images by a factor from a specified range.
Color jitter	Brightness, contrast, saturation, and hue of the image varied a random amount within a specified boundary (+/-) for each image.
Occlusions	Given a probability of occurence, occlusions consisting of two 35x35 pixel black squares are added to the image and its mask at random locations.
Gaussian blur	A kernel of a specified size to include occlusions.
Horizontal and Vertical Flips	For each of the 138 real car photos in the training set, a horizontal and vertical flip have been applied and added to the dataset.

Table 2. Augmentation methods used to expand the dataset

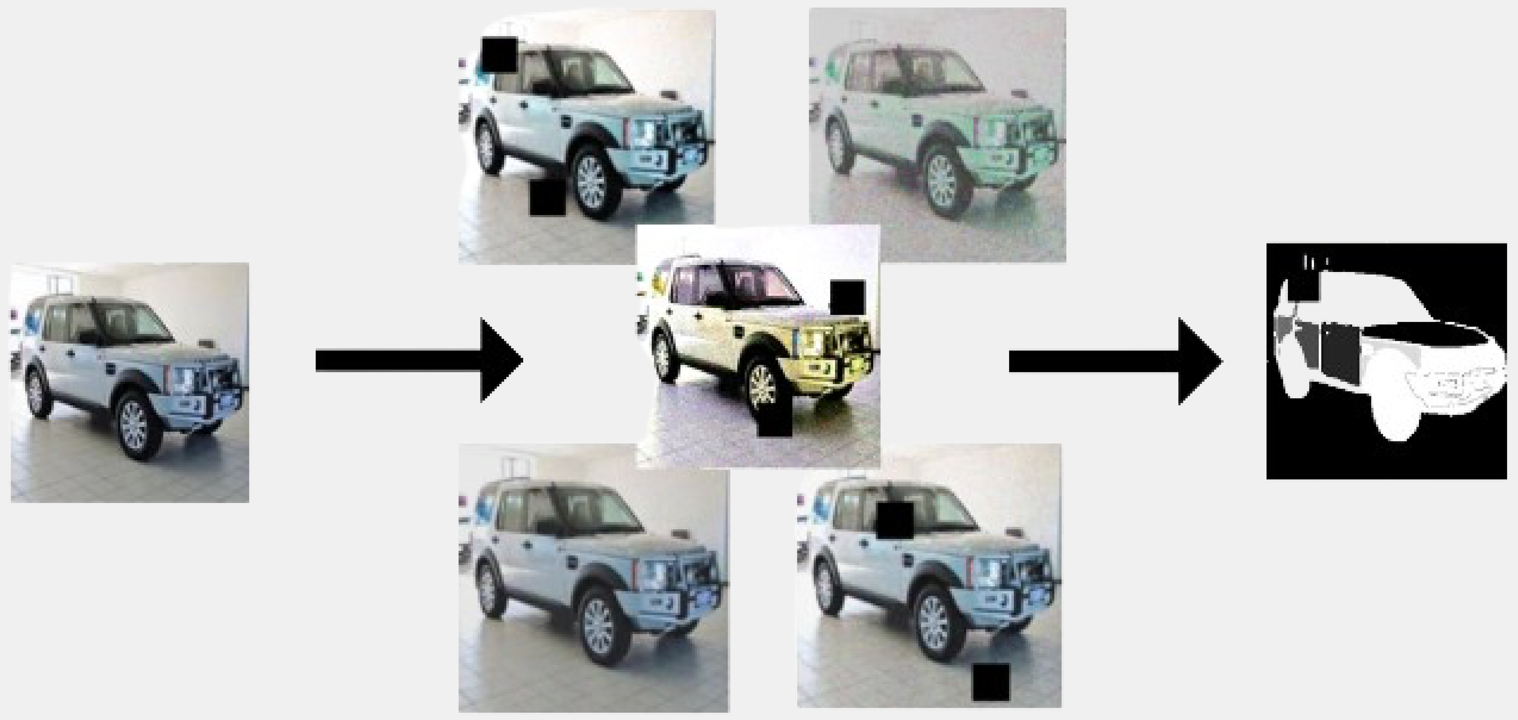


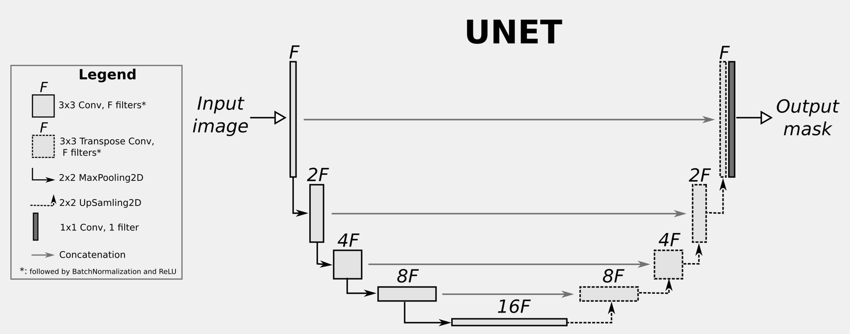
Figure 3. Augmented image along with its corresponding mask

References

[1] Kiprono Elijah Koech.
Cross-entropy loss function, Jul 2022.

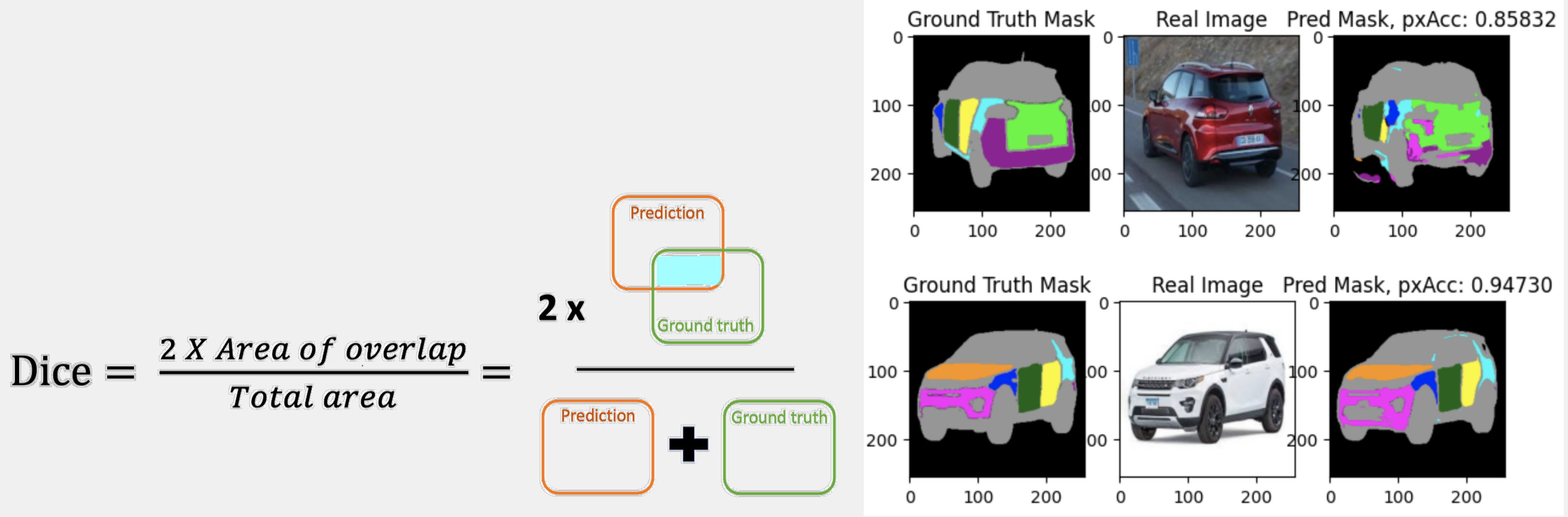
[2] Claude E. Shannon.
A mathematical theory of communication.
Bell System Technical Journal, 27(3):379–423, 1948.

Model architecture



- **Unet+ResNet** has the ability to capture more detailed visual features as it uses ResNet-50's advanced feature extraction capabilities as its encoder, while **Unet** has self-contained convolutional layers.
- **Unet+ResNet** learns from ResNet's pretrained weights, providing a more detailed understanding of complicated image features for improved segmentation accuracy, whereas standard **Unet** learns features only from the given training data.

Metrics



Dice coefficient

Mean pixel difference

Figure 4. Metrics used for performance assessment

$$L_{CE} + L_{DL} = \left(-\sum_{i=1}^{n} t_i \cdot \log(p_i) \right) + \frac{1}{n} \sum_{i=1}^{n} \left(1 - \frac{2 \cdot \sum t_i \cdot p_i}{\sum t_i + \sum p_i + \epsilon} \right) \quad (1)$$

Expression for the sum of Cross Entropy Loss and Dice loss is given above. ti denotes the truth label, n denotes the number of classes pi is the softmax probability of ith class. The dice loss and cross entropy loss are utilized together in the training loop by taking their sum.

Experiments and Results

EXP.	DATA AUGMENTATION	MODEL	EPOCHS	DICE COEFFICIENT	PIXEL ACCURACY
0	0	U-NET	40	0.8844	0.8839
1	1	U-NET	40	0.8822	0.8802
3	2	U-NET	40	0.8788	0.8794
4	3	U-NET	40	0.8709	0.8728
5	4	U-NET	40	0.8876	0.8872
6	0	U-NET+RESNET	40	0.8980	0.8978
7	1	U-NET+RESNET	40	0.8891	0.8921
8	2	U-NET+RESNET	40	0.8685	0.8733
9	3	U-NET+RESNET	40	0.8696	0.8719
10	4	U-NET+RESNET	40	0.8990	0.9005
11	N/A	U-NET	15-40	0.5877	0.57794

Figure 5. Experiment results

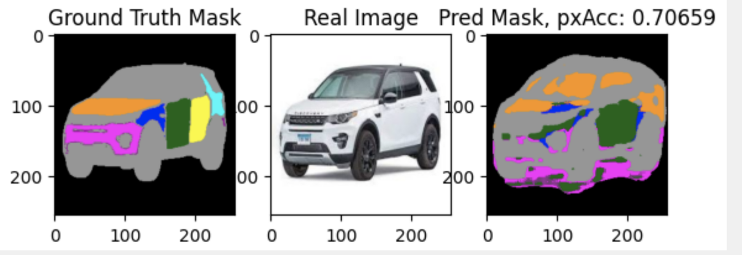


Figure 6. Prediction before optimizations

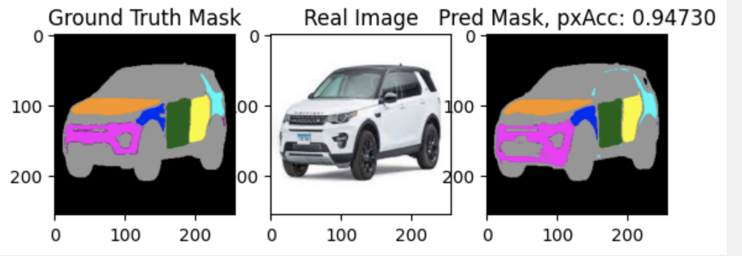


Figure 7. Prediction after optimizations