Deloitte.

Segmentation of car parts

02456 Deep Learning 2023

DTU

Introduction

The problem presented to us is that of car part segmentation in real world images. This can be used in various applications such as insurance claim automation.

The dataset obtained for this task contains:

- 2001 images of a synthetic orange car on a mostly orange background
- 834 images of a synthetic black car on a solid black background
- 168 photos of real cars. Of which 30 are reserved for testing.
- 10.000 photos of backgrounds without cars

Each of the car images has a segmentation with 10 different classes, including a background class and a “rest of car” class.

Fig 1: Examples of each of the 3 car datasets and empty backgrounds

Data processing

We lack real data due to the small size of the photo dataset, to make up for this we used various data augmentation techniques:

- Added random backgrounds to all synthetic images and randomly swap backgrounds in real photos
- Random translations to the cars in the synthetic images
- Random rotations to cars (-30 - 30 deg)
- Random zoom to cars (0.7x - 1.5x)

We also altered the proportion of each of the 3 groups of images in the training dataset. The final dataset used 5x the photo dataset and 400 images from each of the other datasets.

Fig 2: Cars from Fig1. after data augmentation. Notice that the 3rd car can also keep the original background

Models

U-net [1]

U-Net is a convolutional neural network architecture designed for semantic image segmentation, involving convolutional and pooling layers for context extraction.

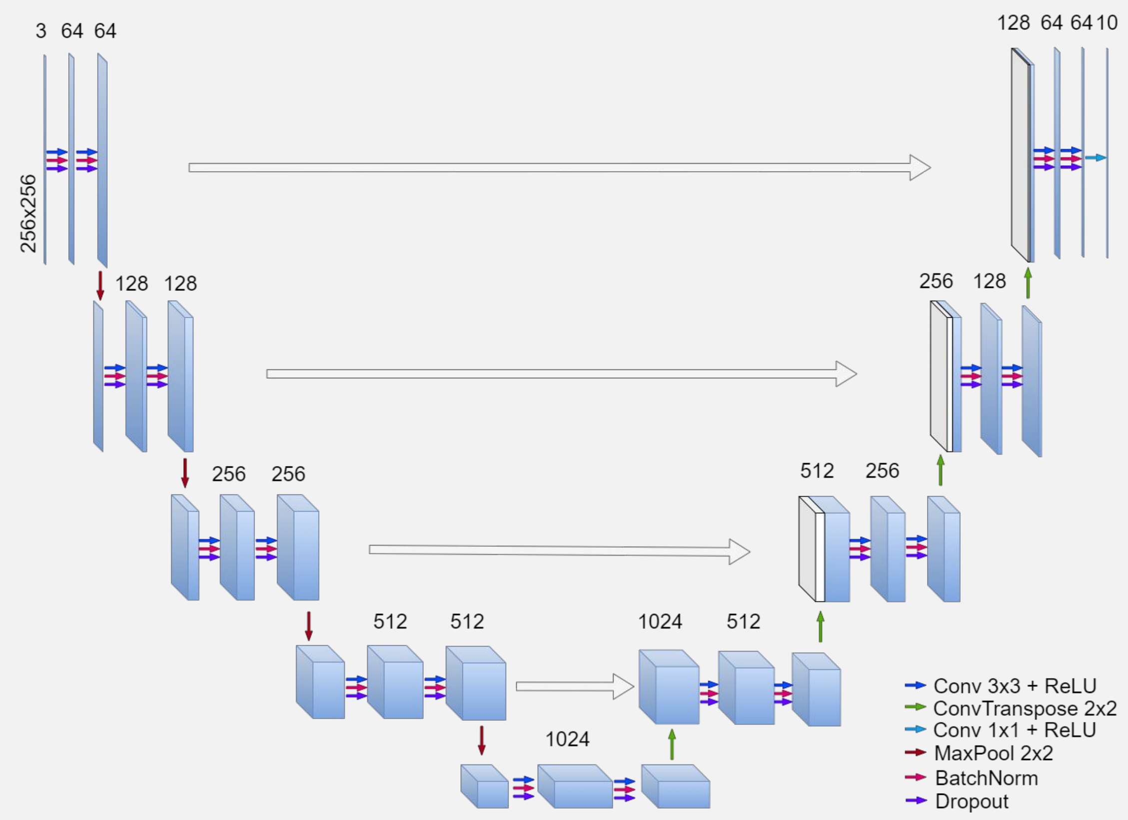

Fig 3: Our UNet architecture. With a Batchnorm layer and Dropout after every convolution.

Pix2pix [2]

The architecture of the Pix2Pix consists of two parts: the generator (UNet) and the discriminator (PatchGAN). The PatchGAN classifies small regions of images as real or fake and provides feedback to the generator based on which the model learns how to generate more realistic segmentations until they are indistinguishable from real ones.

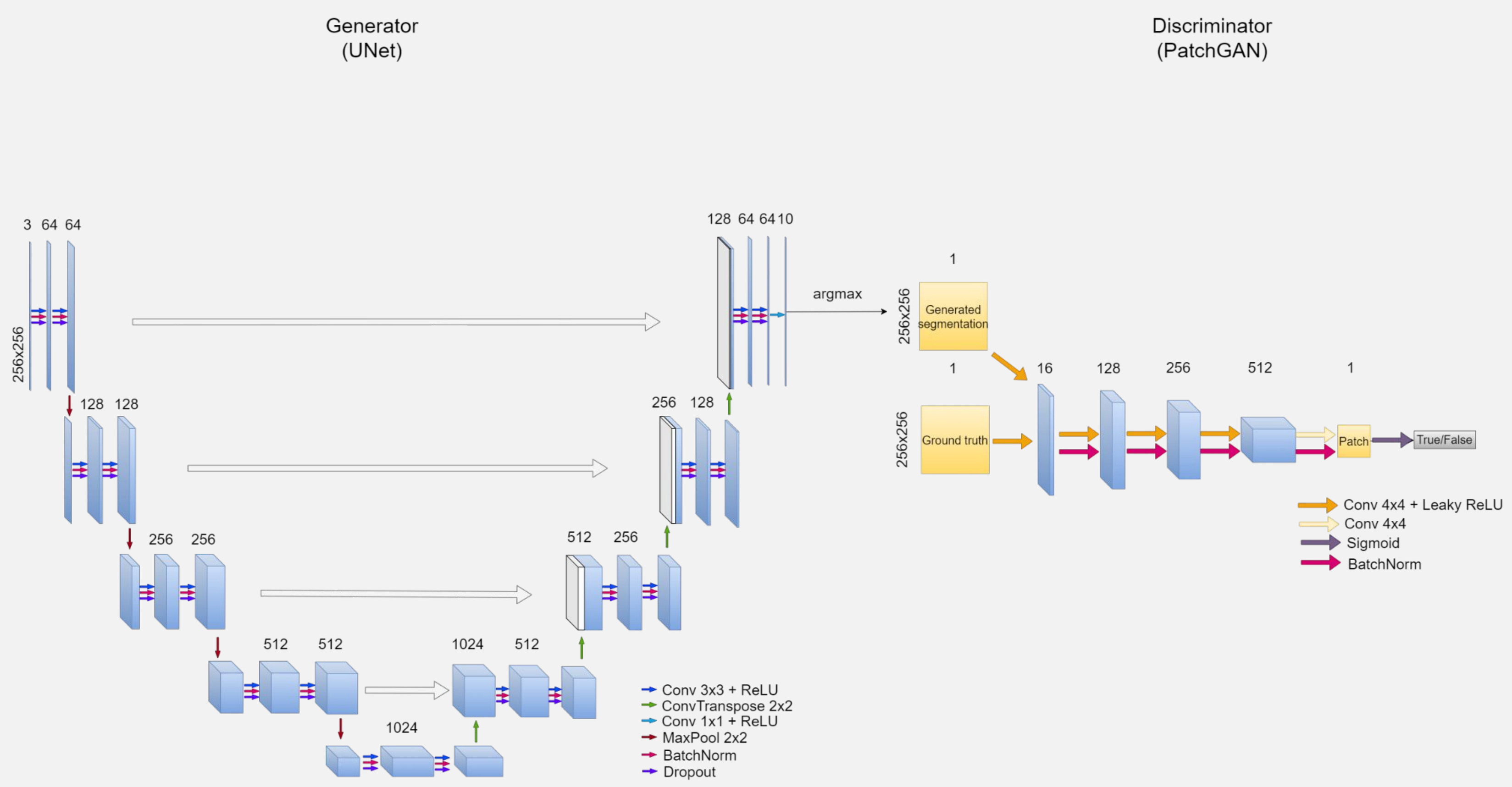

Fig 4: Our Pix2Pix architecture, using our UNet as a Generator and PatchGAN as a Discriminator.

Results

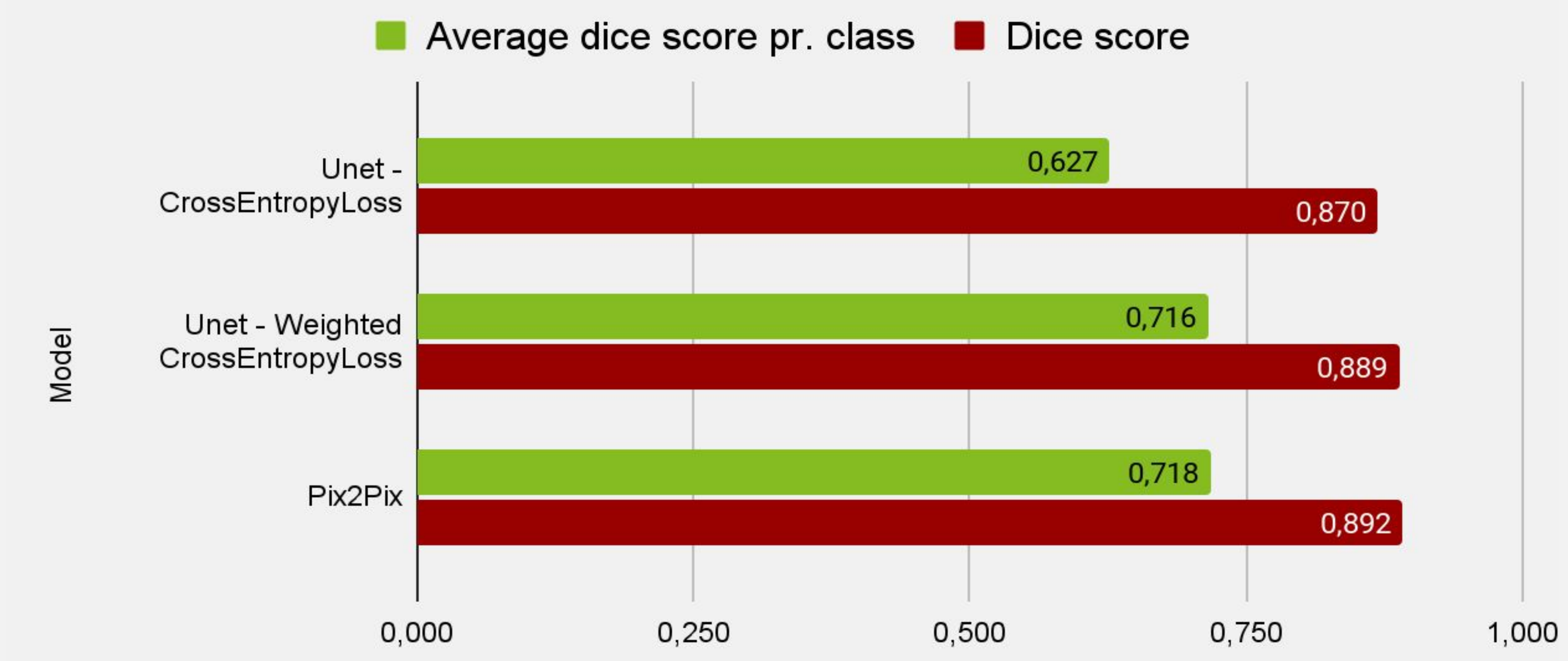

Fig 5: Performance scores of the different models and loss functions tested.

Prediction examples

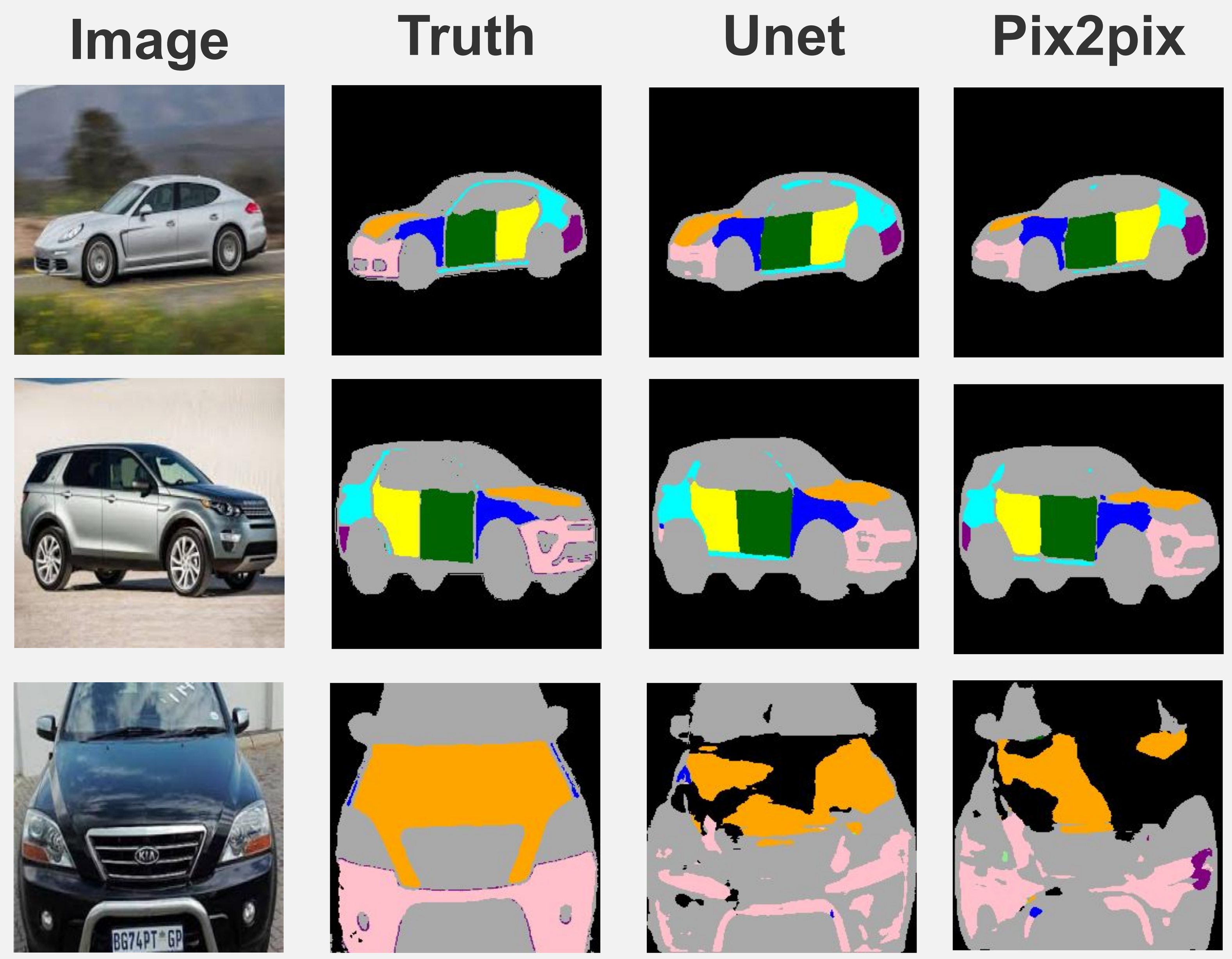

Fig 6: Base truth and outputs of the best performing model from each architecture on 3 different inputs

References

[1] Olaf Ronneberger et al., 2015, *U-Net: Convolutional Networks for Biomedical Image Segmentation*, CoRR

[2] Phillip Isola et al., 2016, *Image-to-Image Translation with Conditional Adversarial Networks*, CoRR

Aleksandra Dominika Ruminska (s232953), Christopher Hardi Nielsen (s194638), Pere Ginebra Solanellas (s223407), Xiaoyu Yan (s232161)