Deep learning - Homework 2

Luka Šveigl (63200301)

1 Introduction

This report outlines our solutions of the three tasks which were part of the second assignment of our Deep learning course. The three tasks were implementation of a Residual neural network for image classification, implementation of a Full convolutional neural network and U-Net for image segmentation, and an adaptation of said U-Net for the task of colorization. In this report we present our implementations, before reporting on the results of the tasks and concluding the report with a discussion.

2 Methodology

2.1 Backbone architecture

The first task of this assignment was the task of image classification, where our neural network had to classify an image into one of 400 different classes. We used a bird dataset which contains 58.400 images in the training set.

In an attempt to create a deep neural network suitable for the task of image classification, we implemented a shallow 18-layer ResNet as described in the original paper [1]. The structure of such a neural network is presented in Figure 1. We trained the model on 20 epochs, using Cross-Entropy loss and the ADAM optimizer with a learning rate of $\gamma = 0.001$ and a linear learning rate scheduler.

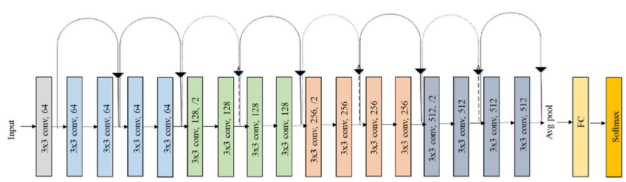

Figure 1: The ResNet18 architecture [3].

2.2 Semantic segmentation

The second task of this assignment was the task of image segmentation, which is the task of partitioning an image into different regions based on their semantic meaning. We used a part of the Lyft self-driving dataset. As part of this task we implemented a Fully convolutional neural network, as described in the paper [2] and a U-Net, described in the paper [4].

Both the FCN and the U-Net have been trained on 10 epochs using the Cross-Entropy loss function and the ADAM optimizer with a learning rate of $\gamma = 0.0001$.

FCN

To implement a Fully convolutional neural network, we adapted our implementation of the classification neural network described in Section 2.1. We modified our implementation by replacing the average pooling layer and the fully connected layer with a segmentation head consisting of a 1×1 convolutional layer, followed by a batch normalization layer, ReLU activation layer, another 1×1 convolutional layer and a bilinear upsampling layer.

U-Net

To implement the U-Net, we followed the architecture of the original paper [4], presented in Figure 2. The U-Net consists of a contracting and expanding path, which are connected by a bottleneck layer. The contracting path consists of multiple serially connected convolutional and pooling layers that reduce the spatial dimensions of their input while increasing the number of output channels. The expanding path consists of several deconvolutional layers, that up-sample their inputs, thus recovering the spatial dimensions of the original image. Skip connections are used to recover the finer details that are lost during the down-sampling process. All convolutional layers of our implementation also include a padding of 1 in order to not reduce the dimensions of the output image.

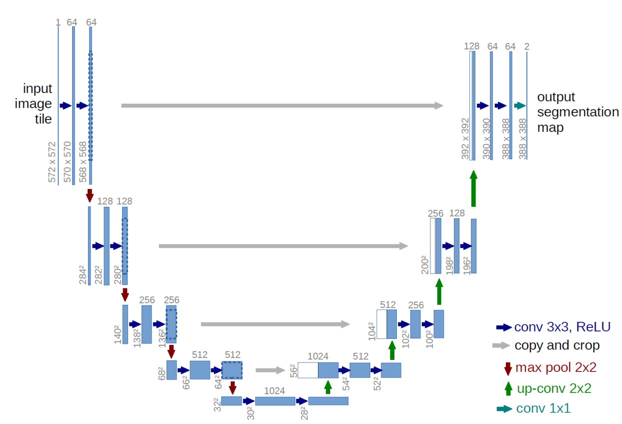

Figure 2: The U-Net architecture [4].

2.3 Colorization

The final task of this assignment was the task of image colorization, for which we had to modify the U-Net implementation mentioned in Section 2.2 for the purpose of coloring grayscale images.

The only modifications to the U-Net was the replacement of the output layer from a convolutional layer with 13 out channels (image segmentation) with a convolutional layer with 3 out channels (image colorization - RGB). Additionally, we implemented a second variant of such a network, where we removed the skip connections between the contracting and expanding paths.

The U-Net implementations were trained on 5 epochs using the Cross-Entropy loss function and the ADAM optimizer with a learning rate of $\gamma = 0.0001$.

3 Results

3.1 Backbone architecture

The ResNet implementation described in Section 2.1 reaches a classification accuracy of 90% and test loss of 0.0656 on the evaluation images.

3.2 Semantic segmentation

To evaluate our networks on the task of image segmentation we used the mean Intersect-over-Union metric, where the Intersect-over-Union is defined by the equation

$$\frac{TP}{TP + FP + FN}.$$

We evaluated each class in each prediction in terms of TP, FP and FN, accumulating the values across all predictions, before calculating the IoU scores for each class. The final mean IoU score is the average of IoU scores of all classes.

The Fully convolutional network achieved a mean IoU score of 0.458, while the U-Net achieved a mean IoU score of 0.478. From the images in Figures 3 and 4 we can observe that the resulting segmentation of the U-Net looks much better and is much more crisp, which is likely due to the skip connections which allow the U-Net to preserve low-level details of the image.

3.3 Colorization

To evaluate our implementations of colorization U-Nets we did not employ any objective metrics. Instead, the results are presented in Figures 5 and 6. We can observe that the U-Net with skip connections produces fairly good and realistic results, while failing to apply more ”unexpected” colors, such as the red of the morning sun. When comparing the U-Net with skip connections to the U-Net without skip connections, we note that the U-Net without skip connections still applies the correct and realistic colors, while also failing to apply more ”unexpected” colors, however the resulting images are extremely blurred. This phenomenon occurs due to the lack of skip connections in the U-Net implementation, which serve in transferring low-level features, which would otherwise be lost during the contracting path, to the output.

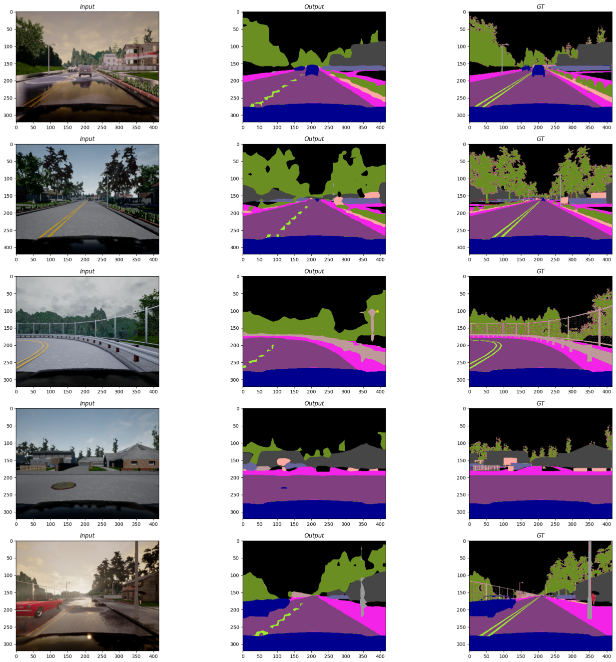

Figure 3: Results of the FCN segmentation.

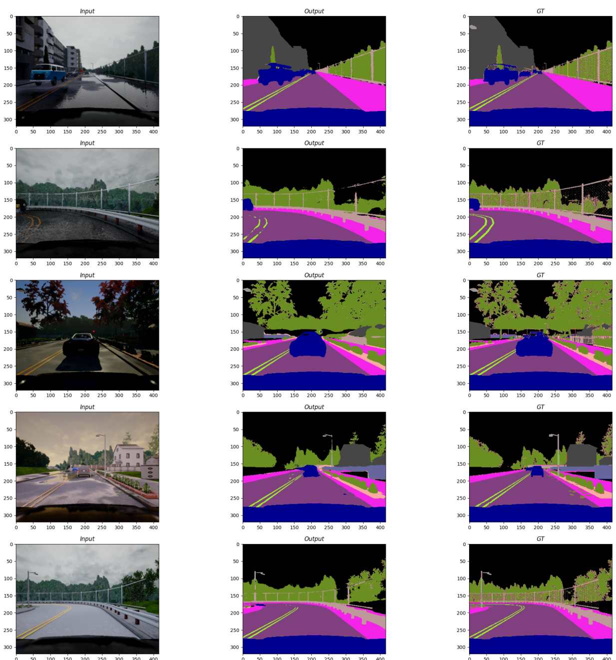

Figure 4: Results of the U-Net segmentation.

4 Discussion

Our assignment was to implement various convolutional neural networks for different image tasks, namely classification, segmentation and colorization. We began by implementing a ResNet with 18 layers for image classification, with which we were able to achieve a 90% classification accuracy on our test set. As part of the second task (image segmentation) we mod-

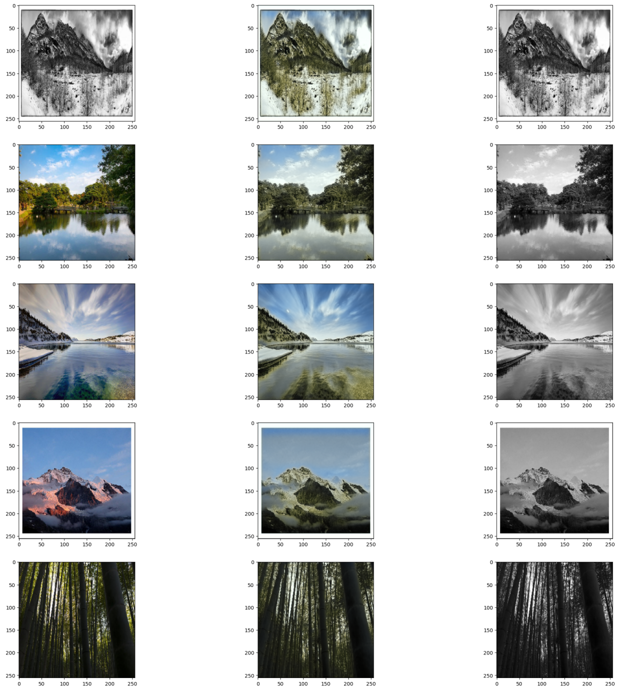

Figure 5: U-Net colorization with skip connections.

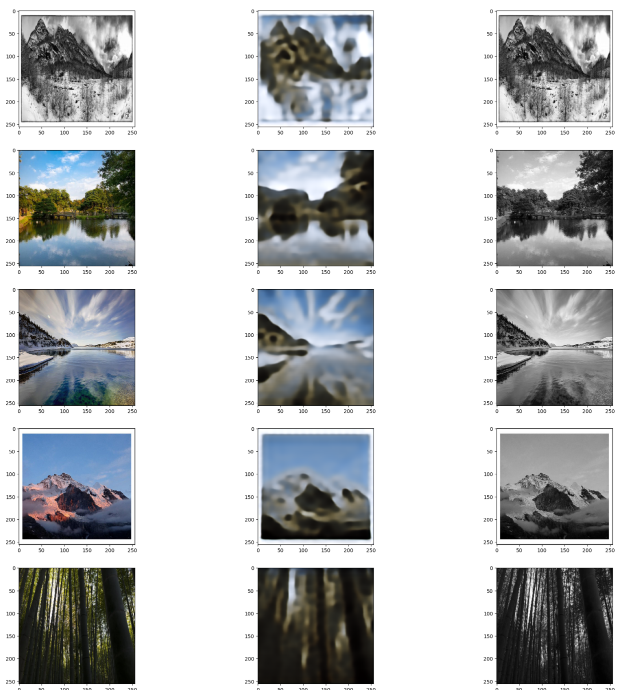

Figure 6: U-Net colorization without skip connections.

ified our ResNet implementation into a Fully convolutional network and implemented a U-Net for image segmentation, with which we achieved mIoU scores of 0.458 and 0.478 respectively. As part of the last task we modified our U-Net implementation for the task of image colorization, training 2 models: one with skip connections and one without skip connections. This enabled us to demonstrate the importance of skip connections in preserving the low-level details of the input.

References

[1] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[2] Jonathan Long, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015, pp. 3431–3440. DOI: 10.1109/CVPR.2015.7298965.

[3] Farheen Ramzan et al. "A Deep Learning Approach for Automated Diagnosis and Multi-Class Classification of Alzheimer's Disease Stages Using Resting-State fMRI and Residual Neural Networks". In: *Journal of Medical Systems* 44 (Dec. 2019). DOI: 10.1007/s10916-019-1475-2.

[4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation." In: *CoRR* abs/1505.04597 (2015). URL: http://dblp.uni-trier.de/db/journals/corr/corr1505.html#RonnebergerFB15.