Deep learning homework 2

Matej Miočić (63180206)

1 Introduction

In this homework we implemented various convolutional neural networks (CNN). First we implemented a shallow 18-layer ResNet [1] for image classification. Then we implemented a fully convolutional network (FCN) [2] and a more SOTA approach - U-Net [3] for image segmentation. Finally we adapted the U-Net implementation for image colorization.

2 Experiments

2.1 Image classification

Image classification using CNNs involves training a deep neural network to recognize and classify images based on their visual features. We implemented a shallow 18-layer ResNet as described in the original paper (see Figure 1). We used a bird dataset which contains 58,400 images in the training set. The task is to classify each image into one of 400 different classes of birds. We use cross-entropy loss for training. For such a task an 18-layer implementation was sufficient enough to obtain 90% classification accuracy on test set. We used the ADAM optimizer with starting learning rate: $\gamma = 0.001$ for 20 epochs. We also added a linear learning rate scheduler.

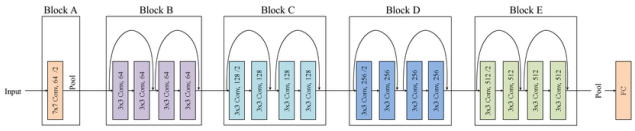

Figure 1: ResNet with 18 layers depicted in five blocks. [4]

2.2 Semantic segmentation

Semantic segmentation is the task of partitioning an image into different regions based on their semantic meaning. We first implemented a fully convolutional network and compared it to U-Net. This time we used part of Lyft self driving dataset which contains 13 classes.

FCN

To implement FCN we took the ResNet with 18 layers and removed the fully connected layer and average pooling layer, and add a segmentation head consisting of a 1x1 convolutional layer, ReLU activation, another 1x1 convolutional layer, and a bilinear upsampling layer.

U-Net

The U-Net architecture (see Figure 2) consists of a contracting path and an expansive path, which are connected by a bottleneck layer. The contracting path consists of a series of convolutional and pooling layers that progressively reduce the spatial dimensions of the feature maps, while increasing the number of channels. The expansive path consists of a series of deconvolutional (transpose convolutional) layers and skip connections that progressively upsample the feature maps and recover the spatial dimensions of the input image.

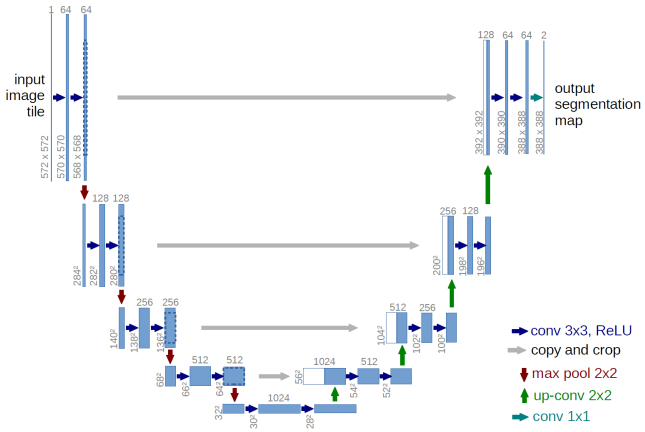

Figure 2: U-Net architecture. [3]

Results

We train the models using cross-entropy loss, and evaluate their performance on the test set using the intersection over union (IOU) metric (1). IOU measures the overlap between the predicted segmentation and the ground truth, and it is calculated for each class by counting the number of true positives (TP), false positives (FP), and false negatives (FN).

$$IOU = \frac{TP}{TP + FP + FN}. \tag{1}$$

We report the mean IOU, which is the average of IOUs for each class. After training for 2 epochs FCN was able to obtain 0.47 IOU, while U-Net achieved 0.55 IOU. The difference seems minimal, but we show a few images from test set

obtained from each network in Figure 3 which shows that the segmentation obtained from U-Net captures the details better; such as leaves on the trees and lines on the road FCN is still able to capture the shape of each class.

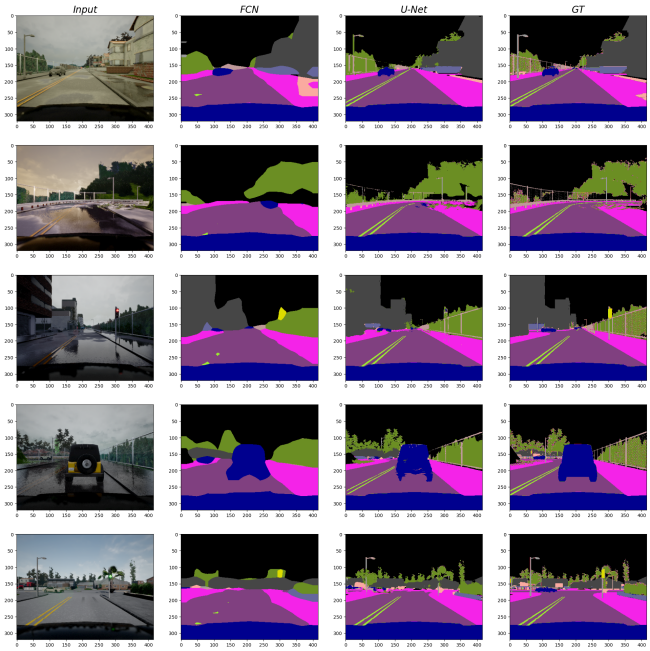

Figure 3: A few examples of segmented test images obtained by FCN (2nd column) and U-Net (3rd column) compared the to ground truth (4th column).

2.3 Image colorization

Image colorization is the process of adding color to a grayscale image. We adapt our U-Net implementation so that now the input is a grayscale image and a colored image as the output. We replace the softmax activation function with a linear activation function that produces the actual color values. We trained 2 versions of networks. One without the U-net skip connections and one with. We trained both for 5 epochs. We show a few examples in Figure 4.

We see that without skip connections, the images become blurred. This is because the U-Net architecture uses skip connections which help in transferring low-level features from the input to the output, while preserving high-level semantic information. This means that without the skip connection the low-level features are lost and images appear blurred.

3 Conclusion

In this homework, we implemented various convolutional neural networks for different image tasks. We started with a ResNet with 18 layers for image classification, achieving 90% accuracy on the bird dataset. Then we implemented a fully convolutional network and U-Net for semantic segmentation, achieving an IOU of 0.47 and 0.55 respectively on a part of the Lyft self-driving dataset. Finally, we adapted the U-Net implementation for image colorization and observed

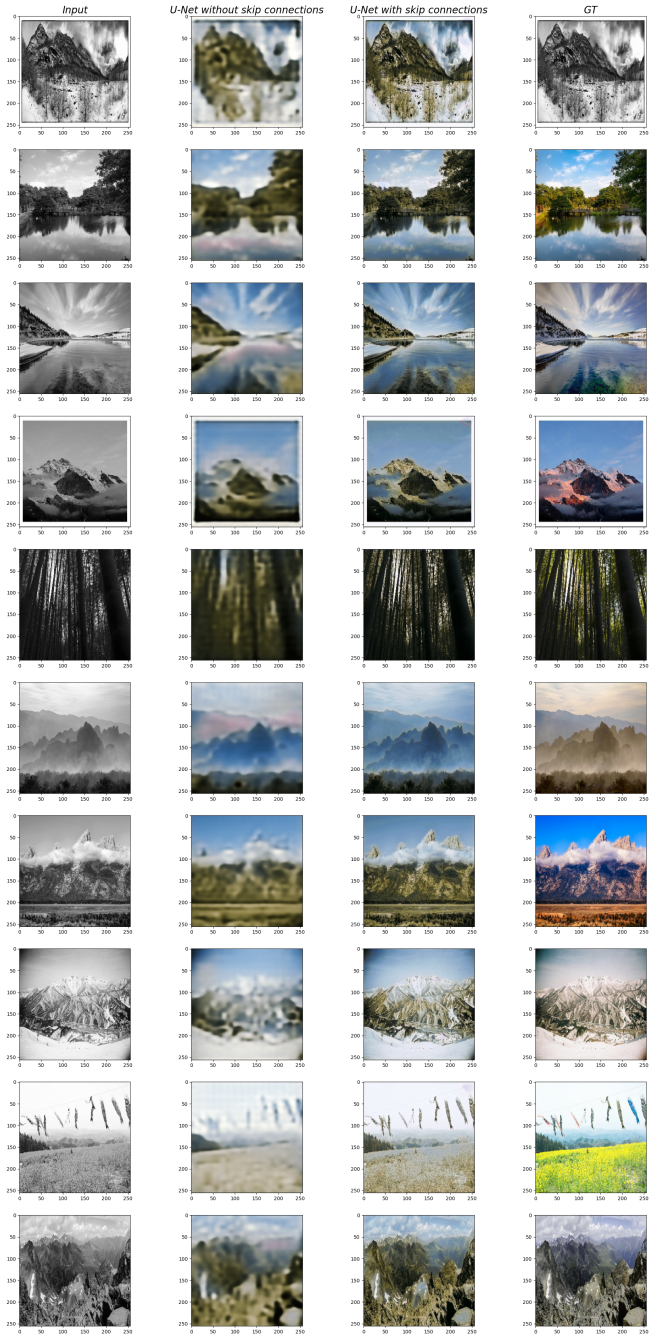

Figure 4: A few examples of colorization of test images obtained by U-Net without skip connections (2nd column) and U-Net with skip connections (3rd column) compared to the ground truth (4th column).

that the use of skip connections is essential to preserve low-level features and obtain high-quality results.

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[2] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation, 2015.

[3] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[4] Sina Ghassemi and Enrico Magli. Convolutional neural networks for on-board cloud screening. *Remote Sensing*, 11:1417, 06 2019.