Deep learning - Homework 2

Stefanela Stevanovic (63220492)

1 Task 1 - Bird image classification with ResNet18

We implemented ResNet18 for the purpose of classifying bird images. The task is to classify each image into one of 400 different classes of birds.

Experiment	Network	Loss Function	Optimizer	LR	Epoch	Cls. Acc.
1	ResNet18	Cross-Entropy	Adam	0.001	10	88%
2	ResNet18	Cross-Entropy	Adam	0.0005	15	90%

Table 1: ResNet18 for bird image classification experiments.

Firstly, we trained our model with a learning rate of 0.001 for 10 epochs, which resulted in an 88% accuracy on the test set. As this accuracy was not satisfactory, we decided to reduce the learning rate to 0.0005 and increase the number of epochs to 15. With these settings, the ResNet model achieved a classification accuracy of 90% on the test set, indicating that the model has a strong capability for generalizing from the training data to unseen data.

2 Task 2 - Image segmentation

For image segmentation, we modified ResNet18 network to include Fully Convolutional Network capabilities - FCN32 and implemented the UNet architecture. To evaluate their performance, we computed the Intersection over Union (IoU) for each class separately and then took the mean over all classes.

Experiment	Network	Loss Function	Optimizer	LR	Epoch	Mean IoU
1	ResNet18	Cross-Entropy	Adam	0.0001	2	38%
2	UNet	Cross-Entropy	Adam	0.0001	2	52%

Table 2: Image segmentation experiments.

Based on the results in Table 2, the UNet architecture outperforms the ResNet18 architecture in terms of mean Intersection over Union (IoU). This difference can be attributed to U-Net's unique structure that incorporates skip connections between its downsampling and upsampling paths, that help preserve spatial information, which is crucial for accurate segmentation.

Figure 1 presents examples of segmentation masks produced by both the FCN-32 and the U-Net networks. The FCN-32 architecture tends to produce more generalized segmentation masks, while the U-Net architecture yields more precise and detailed segmentation masks. This precision is due to its skip connections, which allow the model to utilize both high-level and low-level features during the upsampling process.

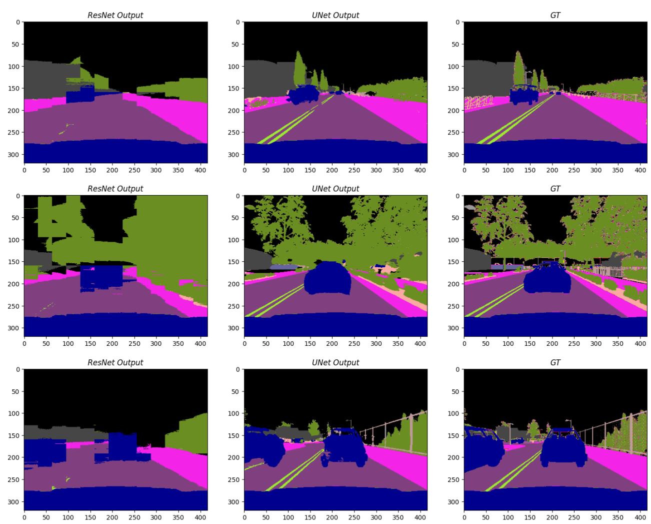

Figure 1: Example segmentation masks from FCN-32 (left) and U-Net (middle) architectures and ground truth (right).

3 Task 3 - Image colorization

For image colorization task we trained two U-Net architectures, with adding or omitting skip connections, to analyze how those changes impact the performance of colorization. In both cases we trained the network for 5 epochs with learning rate of 0.0001. The results are presented on Figure 2.

From Figure 2 we can see that the U-Net with skip connections produces better colorization results compared to its counterpart without skip connections. Skip connections allow information from early layers to be directly concatenated with later layers during upsampling, enabling the model to preserving important features from the input image. Omitting skip connections removes this direct connection and as a result, the output images from UNet without skip connections (modified) look blurred and lack texture present in the input grayscale image.

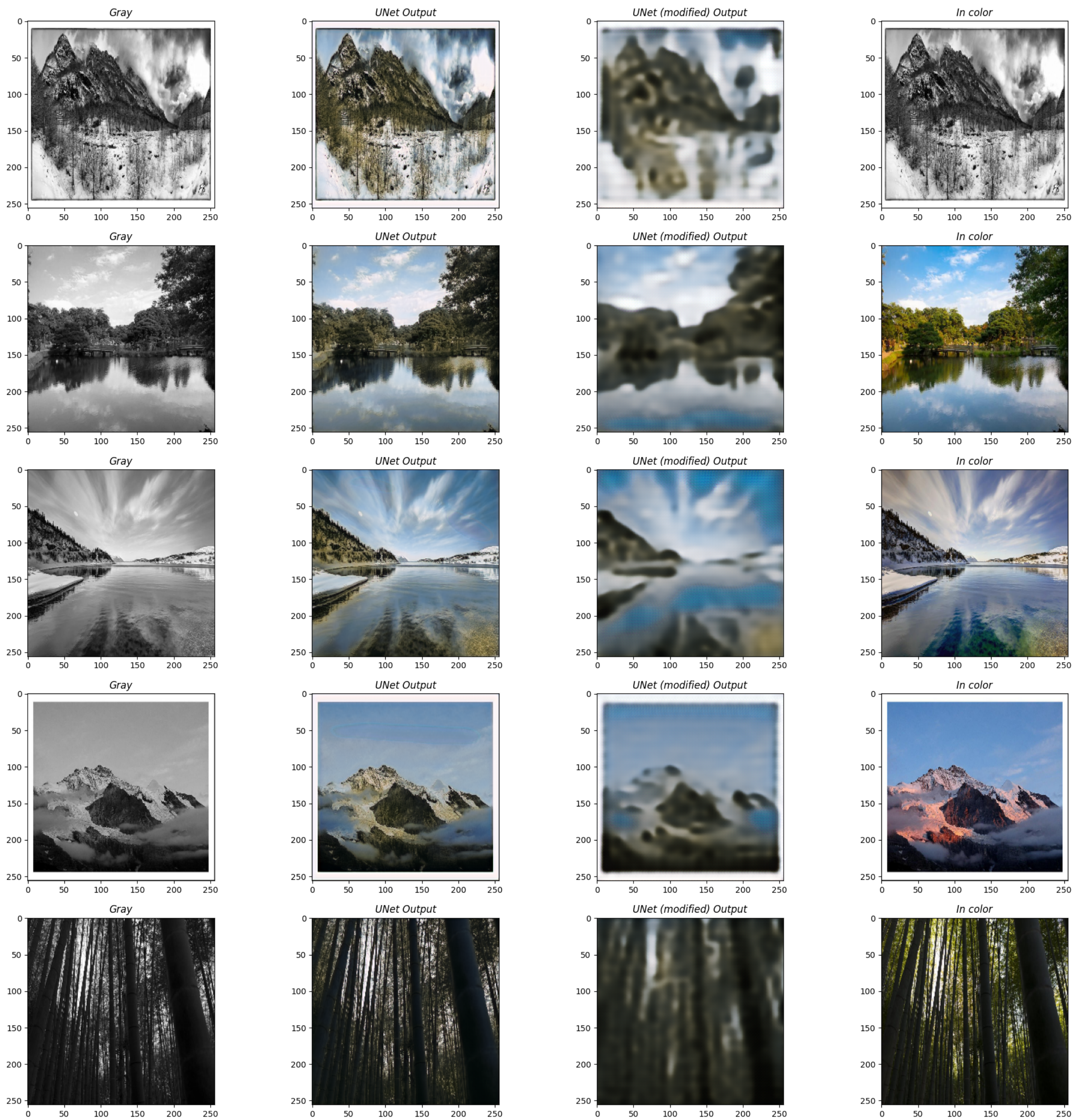

Figure 2: Image colorization results from UNet with skip connections and U-net without skip connections (UNet modified).