Neural Networks and Deep Learning (Khordad 1401)

Assignment #3: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation

Due date: 3th Tir 1401

Image segmentation is the division of an image into regions or categories, which correspond to different objects or parts of objects. After segmentation, the output is a region or a structure that collectively covers the entire image. These regions have similar characteristics including colors, texture, or intensity. Image segmentation is important because it can enhance the analysis of images with more granularity. It extracts the objects of interest, for further processing such as description or recognition.

Image segmentation works by using encoders and decoders. Encoders take in input, which is a raw image, it then extracts features, and finally, decoders generate an output which is an image with segments.
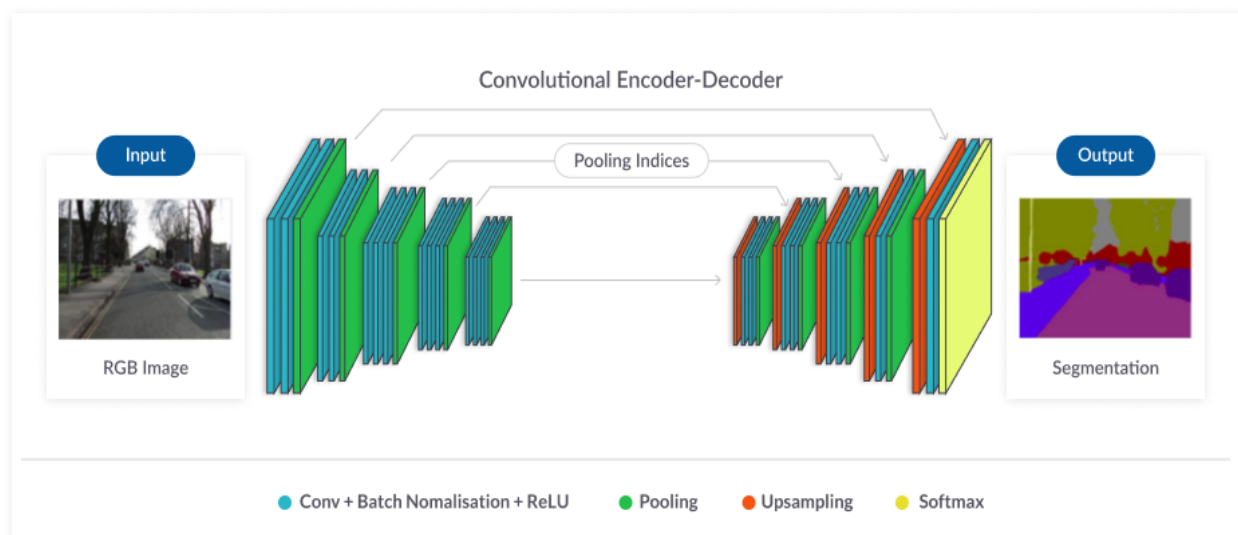

Figure1: The basic model of an image segmentation

The most common architectures in image segmentation are U-Net, FastFCN, Deeplab, Mask R-CNN, etc. Your job is to use U-Net architecture with convolutional layers to build a model for the segmentation task.

Dataset:

Cityscapes data contains labeled videos taken from vehicles driven in Germany. The dataset has images from the original videos, and the semantic segmentation labels are shown in images alongside the original image.

This dataset has 1500 training image files and 200 test image files. Each image file is 256x512 pixels, and each file is a composite with the original photo on the left half of the image, alongside the labeled (mask) image (output of semantic segmentation) on the right half.

- **The Dataset is available here.**

Your Task:

➢ Divide images into two groups, real images and mask images, at 256 pixels on the x axis of each image.

➢ Implement U-Net architecture [1] and train your U-Net model with train images.

➢ Evaluate your model with test images and compare the predicted outputs with mask images. For example:
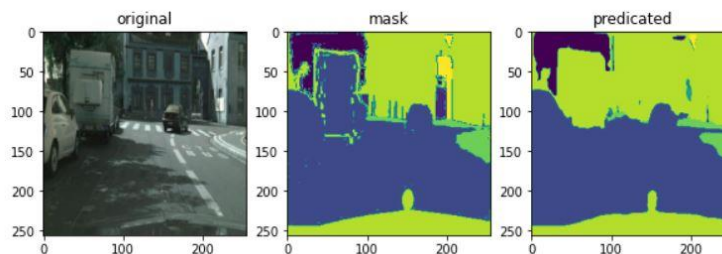
➢ Prepare a report of the network description and also include examples of compared outputs on test images.

---

[1] U-Net architecture is a deep learning image segmentation architecture introduced by Olaf Ronneberger, Philipp Fischer, and Thomas Brox in 2015. Its U shape design consists of two parts. The left side is known as the contracting path or encoder path, where repeated typical convolutions are applied followed by ReLU and max pooling operations. The right side is known as the expansive path which has transposed 2D convolutional layers where the up sampling technique is performed.

Notes:
- ➢ **Pay extra attention to the due date. It will not extend.**
- ➢ **Be advised that submissions after the deadline would not grade.**
- ➢ **Prepare your full report in __PDF__ format and include the figures and results.**
- ➢ **Feel free to use any predefined functions.**
- ➢ **Email your files as a folder in this format (HW#_student#_name_family.zip).**
  - o **Email:** soroushmehrpou@gmail.com