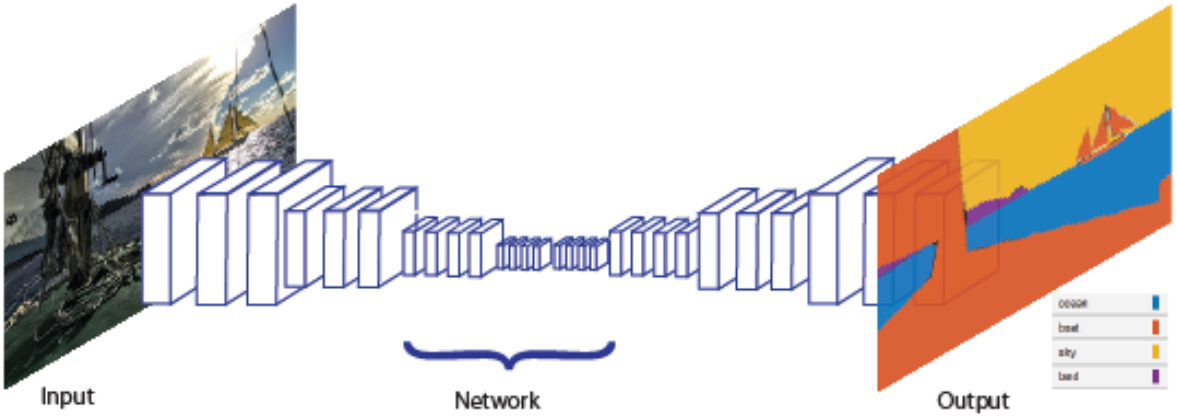
Amir Sotoodeh
CS4990: Machine Learning
11/28/18

Assignment 3: Hair and Skin Segmentation

Introduction

Semantic segmentation has a very practical use in many fields in today's society. Its usage ranges anywhere from biomedical data to autonomous driving purposes. This assignment utilizes FCNN for applications in the field of aesthetics, in which allows potential users to visualize themselves with different hair or skin colors; furthermore, said models could be expanded into the fashion industries for many other practical uses.



The objective of this assignment is to develop a Fully Convolutional Neural Network in which consists of encoder and decoder blocks that will predict the mask for an input and extracts the hair or skin of the individual in the image. The model will have a loss function of binary cross entropy and its accuracy will be based on the produced mask's dice coefficient. The calculation of such metrics will determine whether or not the model's autoencoder/decoder effectively and accurately determines the respective pixels for an image's mask.

Data Used

Initially, I had noticed that the dataset from the kaggle competition was limited to just 2000 total images (combining training and validation); hence I figured that data augmentation or an alternative data collection method would be required. For example, data collection via a mobile application would prove useful to allow the model to become more robust and accurate[1]. However, this would require a platform to obtain this data with an established user-base; hence data augmentation would be the more favorable approach.

Methods and Results

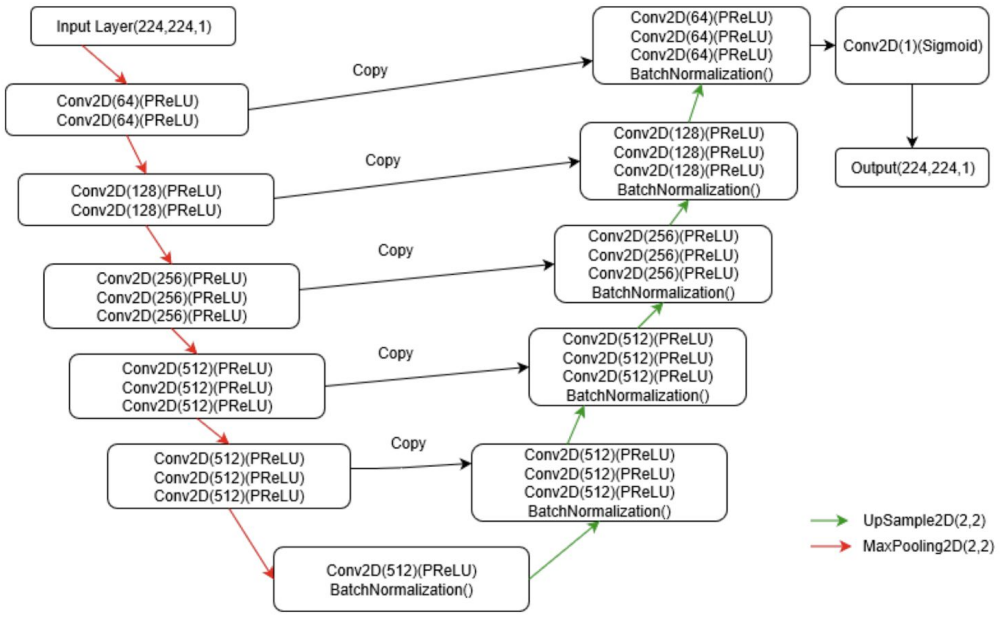
Approach 1:

For hair segmentation, I began my first attempt using a simple U-net architecture in which has an acceptable performance for semantic segmentation. With this architecture, I did not expect a very high dice-coefficient, but rather I expected masks in which cover the general area of where hair is found in an image. The general architecture used was created with guidance from an article in which uses U-Net for biomedical image segmentation[2]. However, this approach yielded a dice coefficient of ~78-79. When observing the produced masks, I found that many of the images with hair colors that blended into the background had many incorrect pixel values. It is clear that there is not enough training data for the model to distinguish between a complex structure like hair and its color along with the background of the image. Hence, I needed a more robust model with pre-trained feature extractors such as the VGG16 architecture. This led me to my next approach in experimenting with a different architecture.

[1] Aarabi, Guo, et al., “Real-time deep hair matting on mobile devices”
[2] Fischer, et al., “U-Net: Convolutional Networks for Biomedical Image Segmentation”

Approach 2:

Following the discovery of the first model's flaws, I took the approach of using a pre-trained autoencoder from the VGG16 architecture[3]. This architecture utilizes the first 13 convolutional layers from VGG16 with weights pre-trained on the imagenet dataset and is followed by UpSampling operations to restore the dimensions of the original input image. Additionally, skip connections are made between the last convolution layers of each block in the convolution layers which are concatenated to the respective Conv2DTranspose layers during the UpSampling process.



With this new architecture, I was able to obtain an improved dice coefficient of 82; however, I found that this model has similar flaws in comparison to the base U-Net model. Given that hair is a complex feature to segment, more data and perhaps a more robust model would be required to obtain a higher dice-coefficient. In order to provide the model more data, I generated 2000 more images in which are horizontally flipped images of the original dataset;

[3] Balakrishna, et al., "Automatic detection of lumen and media in the IVUS images using U-Net with VGG16 Encoder"

this results in 4000 total images in which is split to 3600 training images and 400 validation images. I ran into some difficulties with discerning better augmentation methods given that utilizing a data_generator from keras will modify both the training image and mask. I was unsure whether or not color changes would have a significant impact on the model's accuracy and decided not to proceed with any further data augmentation.

With respect to skin segmentation, I utilized the same VGG-16 U-Net hybrid architecture was able to obtain a dice coefficient of ~95.5; which held first place in the competition as of Monday, November 26 10:34 PM. I expected this model to perform much better for skin segmentation given that skin is much easier to distinguish and differentiate it from the image. To train the model for both skin and hair, I utilized the Adadelta optimizer with the following parameters:

- Learning rate: 1.0
- $\rho = 0.95$
- $\epsilon = 1e^{-7}$

Future Prospects

I plan to begin researching training with Caffe models as well as experimenting with the FCN8s/FCN32s/E-NET architecture and potentially attempt to utilize Conditional Random Field (CRF) along with a hair-matting extension [1]. I hope to integrate the aforementioned features into a Python Flask server that users a webcam and submit it for my final project if time permits.

Resources

Fischer, et al., *U-Net: Convolutional Networks for Biomedical Image Segmentation.* Retrieved From https://arxiv.org/pdf/1505.04597.pdf

Aarabi, Guo, et al., *Real-time deep hair matting on mobile devices.* Retrieved From https://arxiv.org/pdf/1712.07168.pdf

Balakrishna, et al., *Automatic detection of lumen and media in the IVUS images using U-Net with VGG16 Encoder.* Retrieved From https://arxiv.org/pdf/1806.07554.pdf