Project Milestone: Convolutional Neural Network to Image Segmentation

Felipe Augusto Lima Reis

PUC Minas - Pontifícia Universidade Católica de Minas Gerais

R. Walter Ianni 255 - Bloco L - Belo Horizonte, MG, Brasil

`falreis@sga.pucminas.br`

Abstract

The ABSTRACT is to be in fully-justified italicized text, at the top of the left-hand column, below the author and affiliation information. Use the word "Abstract" as the title, in 12-point Times, boldface type, centered relative to the column, initially capitalized. The abstract is to be in 10-point, single-spaced type. Leave two blank lines after the Abstract, then begin the main text. Look at previous CVPR abstracts to get a feel for style and length.

1. Introduction

Image segmentation refers to the partition of an image into a set of regions to cover it, to represent meaningful areas [5]. The goal is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze [2].

Segmentation has two main objectives: the first one is to decompose the image into parts for further analysis and the second one is to perform a change of representation [5]. Also, segmentation must follow some characteristics to identify regions, as it follows:

- Regions of an image segmentation should be uniform and homogeneous with respect to some characteristic, such as gray level, color, or texture [5];

- Region interiors should be simple and without many small holes [5];

- Adjacent regions of a segmentation should have significantly different values with respect to the characteristic on which they are uniform [5];

- Boundaries of each segment should be smooth, not ragged, and should be spatially accurate [5].

The future paper will evaluate segmentation methods using Deep Neural Networks and compares with classical methods of segmentation, using the superpixels approach.

Also, the paper will evaluate the composition of classical methods with DNN approach, to speed up the training process and become more accurate.

The organization of this paper is as follows. In the next Section we discuss the problem statement. In Section 3 its explained how the will work and the results we expect. Then in Section 4 we present an some preliminary results.

2. Problem Statement

Semantic pixel-wise segmentation is an active topic of research [4]. Before the use of deep neural networks, the best performing methods mostly was made using hand engineered features [4].

The success of deep convolutional neural networks for object classification led researchers to use these technology to learn new capabilities, such as segmentation [4].

In future paper it will evaluate some Deep Neural Network to segmentation, as SEGNET [4] and U-NET [8], and compare to classical methods, like SLIC (Simple Linear Iterative Clustering) [1] and EGB (Ecient Graph-Based Image Segmentation) [6]. For this, it will be used Berkeley Segmentation Data Set 500 (BSDS500) [3].

Berkeley Segmentation Data Set contains 500 natural images and its respectives ground-truths, annotated by humans [3]. The images are explicitly separated into disjoint train, validation and test subsets [3].

To evaluate the quality of the segmentation methods, the results will be evaluated with BSDS500 benchmarking tool, provided with the Dataset [3]. BSDS500 dataset uses Precision and Recall Method to evaluate the results [3].

This work expects better performance of Deep Neural Network when compared with classical methods (SLIC and EGB). The results must be more precise, but with time and space complexity bigger then classical algorithms.

3. Technical Approach

To provide the goals explained in Section 2, it will be used Convolutional Neural Networks provided by the literature.

3.1. Deep Neural Networks

The project will use two different Neural Networks to provide segmentation, as its follows in the nexts subsections.

3.1.1 SEGNET

SEGNET is a deep encoder-decoder architecture for multi-class pixelwise segmentation [4]. The SEGNET architecture consists of a sequence of non-linear processing layers (encoders) and a corresponding set of decoders followed by a pixelwise classifier [4] [10]. Typically, each encoder consists of one or more convolutional layers with batch normalisation and a ReLU non-linearity, followed by non-overlapping maxpooling and sub-sampling [4] [10]. The sparse encoding due to the pooling process is upsampled in the decoder using the maxpooling indices in the encoding sequence [4] [10]. Figure 1 presents the representation of SEGNET's architecture.
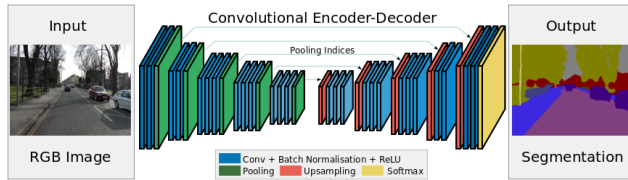

Figure 1. SEGNET architecture. *Image adapted from SEGNET project website* [10] [4]

3.1.2 U-NET

U-NET is a Convolutional Networks for Biomedical Image Segmentation [8] [7]. Although U-NET was developed to biomedical image segmentation, its architecture can be trained to segment other types of image. In this project, we will use U-NET to classify images from BSDS500.

U-NET architecture consists of the repeated application of two 3x3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for downsampling [8]. At each downsampling doubles the number of feature channels [8]. Every step in the expansive path consists of an upsampling of the feature map followed by a 2x2 convolution, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3x3 convolutions, each followed by a ReLU [8]. At the final layer a 1x1 convolution is used to map each 64-component feature vector to the desired number of classes. In total the network has 23 convolutional layers [8]. Figure 2 presents the representation of U-NET's architecture.

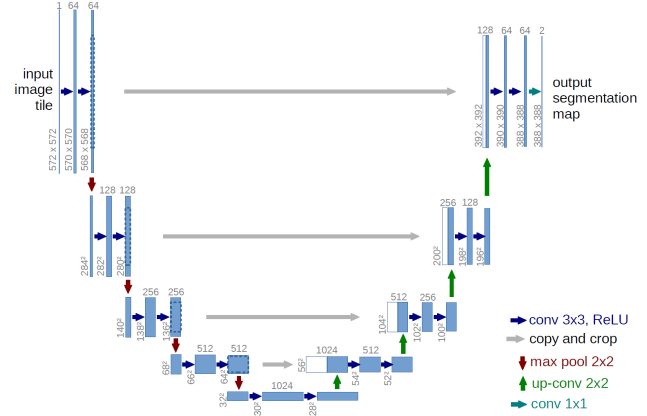### 3.2. Transfer Learning

[?]

Figure 2. U-NET architecture. *Image adapted from U-NET project website* [7] [8]

[9] .

3.3. Data Augmentation

4. Preliminary Results

References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Ssstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, Nov 2012.

[2] S. A. Ahmed, S. Dey, and K. K. Sarma. Image texture classification using artificial neural network (ann). In *2011 2nd National Conference on Emerging Trends and Applications in Computer Science*, pages 1–4, March 2011.

[3] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(5):898–916, 5 2011.

[4] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *CoRR*, abs/1511.00561, 2015.

[5] D. Domnguez and R. R. Morales. *Image Segmentation: Advances*, volume 1. 2016.

[6] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, Sep 2004.

[7] V. P. Recognition and F. o. E. Image Processing, Dept. of Computer Science. U-net: Convolutional networks for biomedical image segmentation, 2018.

[8] O. Ronneberger, P.Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241. Springer, 2015. (available on arXiv:1505.04597 [cs.CV]).

[9] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

[10] C. Vision and U. Robotics Group at the University of Cambridge. Segnet, 2018.