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 [8]. The goal is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze [3].

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 [8]. 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 [8];
- Region interiors should be simple and without many small holes [8];
- Adjacent regions of a segmentation should have significantly different values with respect to the characteristic on which they are uniform [8];
- Boundaries of each segment should be smooth, not ragged, and should be spatially accurate [8].

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 [5]. Before the use of deep neural networks, the best performing methods mostly was made using hand engineered features [5].

The success of deep convolutional neural networks for object classification led researchers to use these technology to learn new capabilities, such as segmentation [5].

In future paper it will evaluate some Deep Neural Network to segmentation, as SEGNET [5] and U-NET [14], and compare to classical methods, like SLIC (Simple Linear Iterative Clustering) [2] and EGB (Ecient Graph-Based Image Segmentation) [10]. For this, it will be used Berkeley Segmentation Data Set 500 (BSDS500) [4].

Berkeley Segmentation Data Set contains 500 natural images and its respectives ground-truths, annotated by humans [4]. The images are explicitly separated into disjoint train, validation and test subsets [4].

To evaluate the quality of the segmentation methods, the results will be evaluated with BSDS500 benchmarking tool, provided with the Dataset [4]. BSDS500 dataset uses Precision and Recall Method to evaluate the results [4].

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 [5]. The SEGNET architecture consists of a sequence of non-linear processing layers (encoders) and a corresponding set of decoders followed by a pixelwise classifier [5] [18]. 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 [5] [18]. The sparse encoding due to the pooling process is upsampled in the decoder using the maxpooling indices in the encoding sequence [5] [18]. Figure 1 presents the representation of SEGNET's architecture.

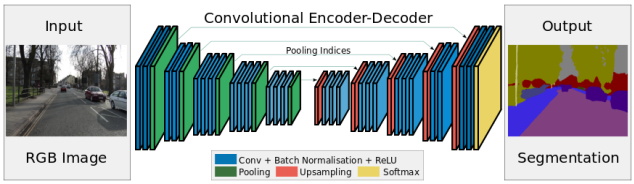

Figure 1. SEGNET architecture. *Image adapted from SEGNET project website* [18] [5]

3.1.2 U-NET

U-NET is a Convolutional Networks for Biomedical Image Segmentation [14] [13]. 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 [14]. At each downsampling doubles the number of feature channels [14]. 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 [14]. 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 [14]. Figure 2 presents the representation of U-NET's architecture.

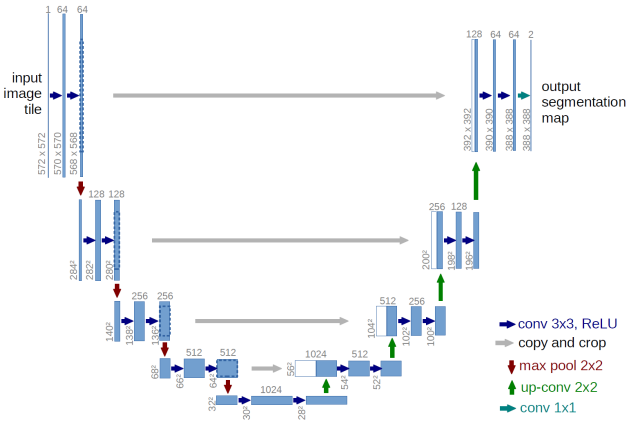

Figure 2. U-NET architecture. *Image adapted from U-NET project website* [13] [14]

3.2. Transfer Learning

Transfer learning is a technique in machine learning that stores knowledge gained while solving one problem, adapt and apply it to a different but related problem. As the grow of neural networks usage, it becomes reasonable to seek out methods that avoid "reinventing the wheel", and instead are able to build on previously trained networks' results [12] [19].

In this work it expected to use transfer learning to speed up the training process. For that, it will be used a pretrained VGG-16 (Very Deep Convolutional Networks for Large-Scale Image Recognition) [16]. The pretrained VGG-16 will be provided by Keras, a Python Deep Learning Library [6]. Keras is a high-level neural networks API of running on top of TensorFlow [1], CNTK [15], or Theano [17] [6].

VGG-16 provided by Keras contains weights pre-trained on ImageNet Dataset [7].

During training, the input to our ConvNets is a fixed-size 224 224 RGB image. The only pre-processing we do is subtracting the mean RGB value, computed on the training set, from each pixel. The image is passed through a stack of convolutional (conv.) layers, where we use filters with a very small receptive field: 3 3 (which is the smallest size to capture the notion of left/right, up/down, center). In one of the configurations we also utilise 1 1 convolution filters, which can be seen as a linear transformation of the input channels (followed by non-linearity). The convolution stride is fixed to 1 pixel; the spatial padding of conv. layer input is such that the spatial resolution is preserved after convolution, i.e. the padding is 1 pixel for 3 3 conv. layers. Spatial pooling is carried out by five max-pooling layers, which follow some of the conv. layers (not all the conv. layers are followed by max-pooling). Max-pooling is performed over a 2 2 pixel window, with stride 2 [16].

3.3. Data Augmentation

Data augmentation consists in a range of transformations that can be applied to dataset to increase the number of data with the target of improving the accuracy and robustness of classifiers [9]. The problem with small datasets is that models trained with them do not generalize well [11].

Data augmentation also can act as a regularizer in preventing overfitting in neural networks and improve performance in imbalanced class problems [20]. According to Wong et al. [20], data augmentation is better to perform in data-space instead of feature-space, as long as label preserving transforms are known [20].

Once BSDS500 contains only 200 images for training and 100 images for validation, the Neural Network may not generalize well and learn enough information from the dataset. Then, it's necessary provide a range of transformation to add some generated images for training and validation.

To provide data augmentation, the images and the ground-truth will be rotated 12 times, 30 degrees each. Also, the images will be flipped and rotated 12 times each. Then, each image will transform into 24 possible images. Then, 200 images for training set will become 4800 training images and the validation set will contains 2400 images. The number of images is not too big, but can help the DNN predict with more accuracy.

4. Preliminary Results

References

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[2] 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.

[3] 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.

[4] 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.

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

[6] F. Chollet et al. Keras. `https://keras.io`, 2015.

[7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

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

[9] A. Fawzi, H. Samulowitz, D. Turaga, and P. Frossard. Adaptive data augmentation for image classification. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 3688–3692, Sept 2016.

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

[11] L. Perez and J. Wang. The effectiveness of data augmentation in image classification using deep learning. *CoRR*, abs/1712.04621, 2017.

[12] L. Y. Pratt. Discriminability-based transfer between neural networks. In *Proceedings of the 5th International Conference on Neural Information Processing Systems*, NIPS'92, pages 204–211, San Francisco, CA, USA, 1992. Morgan Kaufmann Publishers Inc.

[13] V. P. Recognition and F. o. E. Image Processing, Dept. of Computer Science. U-net: Convolutional networks for biomedical image segmentation. `https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/`, 2018.

[14] 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]).

[15] F. Seide and A. Agarwal. Cntk: Microsoft's open-source deep-learning toolkit. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 2135–2135, New York, NY, USA, 2016. ACM.

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

[17] Theano Development Team. Theano: A Python framework for fast computation of mathematical expressions. *arXiv e-prints*, abs/1605.02688, 5 2016.

[18] C. Vision and U. Robotics Group at the University of Cambridge. Segnet. `http://mi.eng.cam.ac.uk/projects/segnet/`, 2018.

[19] K. Weiss, T. M. Khoshgoftaar, and D. Wang. A survey of transfer learning. *Journal of Big Data*, 3(1):9, May 2016.

[20] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell. Understanding data augmentation for classification: when to warp? *CoRR*, abs/1609.08764, 2016.