UDC 004.93

APPROACHES FOR ALGORITHM SELECTION IN SEMANTIC IMAGE SEGMENTATION PROBLEM

Kim A.S.
alkim@nu.edu.kz
Student, Nazarbayev University, Astana
Supervisor – M. Lukac

1 Introduction

Semantic image segmentation is a challenging problem in machine vision and mainly consists of two parts: segmenting images into coherent regions and detecting objects inside of those regions. One may think about this problem as finding a way to assign object labels to meaningful parts of an image. For example, figure 1 shows an image for VOC 2012 dataset [1] with its corresponding ground truth image of semantic segmentation.


(a)

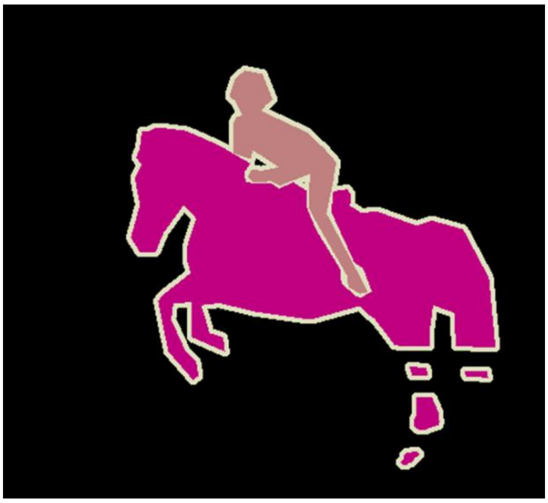
(b)

Figure 1: (a) an image from VOC 2012 dataset; (b) the desired result of semantic segmentation of (a)

Various methods have been used in different algorithms in order to solve the problem of semantic image segmentation. Some approaches are graph-based as in [2]; [3] combines segmentation and recognition, [4] deploys texture and spatial layout features as well as context information, work in [5] is based on probabilistic models, [6, 7] deploy deep learning while [8] view the problem as an image elements classification task.

Algorithm selection is a separate problem in the field of computer science. This approach was first introduced by [9] for the selection of operating system schedulers. In this paper, we are going to apply algorithm selection in the context of image understanding. For semantic image segmentation, algorithm selection is an algorithm that, from the pool of available segmentation algorithms, finds the one that is likely to perform best on a given picture. In our case, to do so, the algorithm selector retrieves features and attributes of an image and applies machine learning techniques to choose the candidate that has most potential to perform well on the image.

It is possible to retrieve a number of various features from an image. In our experiments, such features include brightness levels, sift, gist, FFT, wavelet, Gabor, color, etc. In this paper, for the purpose of distinguishing those features from CNN features, they will be referred to as *normal* features.

In this work, we are going to deploy an artificial neural network (ANN) that will be fed normal features and/or convolutional neural network (CNN) features and some statistical data in the problem of algorithm selection. We are also going to compare the performance of the selector when using a support vector machine to that when an ANN is used. Then the hypothesis that accurate high level attributes are helpful in algorithm selection problem will be tested and justified.

2 Methods and Results

This work is based upon the previous research on algorithm selection by [10] in which the selection method using support vector machines (SVM), Bayesian networks (BN) and high level attributes, was introduced. The features that were used in [10] were only normal features; the attributes were numeric values that were extracted by the MATLAB *regionprops* function that calculates centroids of objects in the picture, major and minor axis lengths, orientations, areas and other properties.

Throughout all of the experiments, we are going to use five of the following algorithms [3], [6], [5], [7] and [8], mainly because the segmentation results of those algorithms are available and the experiments were performed on the same VOC 2012 dataset. For the convenience, in this paper we are going to call those algorithms as CPMC, SDS, ALE, FCN and FFD respectively.

2.1 Accuracy of selection

First, we are going to extract CNN features using Caffe framework [11]. Then we will perform principal component analysis (PCA) algorithm to reduce the dimensions of the obtained features and we will feed the CNN features (with/without normal features) to an ANN to test the accuracy of selection. Then we are going to compare the ANN method’s performance to that of the SVM’s. For a fair comparison, we have also performed the identical experiment with SVM, so that it is clear that the accuracy of the two approaches differs only because of the methods used and not because of the set of features.

The figure 2 illustrates the results obtained through the experiments as described above. The percentage in each cell is the percentage of the number of times the best algorithm was chosen for the segmentation task. The superiority of ANN over CNN is observed with ANN performing more than 10% better than SVM.

	SVM	ANN
CNN features	37.29%	48.48%
all features	37.53%	49.56%

Figure 2: Algorithm selection accuracy for SVM and ANN using different sets of features

2.2 Using statistical data

In [7] an iterative analysis approach was implemented. In their method, a trained SVM is the initial selector and then, if the contradiction is detected in the chosen segmentation image, a new hypothesis is generated and the region containing the contradiction is replaced with a new hypothesis. This process continues until there are no more contradictions or if all of the algorithms have been tried. Hypothesis is represented by an object label; both hypothesis and contradiction are generated by the co-occurrence statistics.

In our work, we deploy a similar approach however we use an ANN instead of an SVM and when a contradiction is found, we use statistics to decide which algorithm might be better suited to perform on that region.

The performance of our algorithm selection approach will be compared with that of the five algorithms that were used in the experiment. The results were analyzed pixel-wise: by the amount of image pixels that were assigned the desired object labels.

The comparison is demonstrated below in figure 3. It can be seen that in general, the algorithm selection approach was slightly better than the best algorithm. It also outperformed all of the available algorithms in 10 categories of objects, namely in the categories of aeroplane, bird, boat, bus, chair, cow, dining table, dog, sofa and train (in figure 3 whose are highlighted in bold). The results are promising and improving the accuracy of selection would result in even better result of segmentation. This leads us to the next section that demonstrates the ways in which the selection accuracy can be improved.

	Used algorithms					
Object class	CPMC	SDS	ALE	FCN	FFD	Algorithm selector
background	83.10%	84.84%	71.51%	91.59%	91.94%	91.91%
aeroplane	64.40%	60.92%	52.16%	82.88%	81.33%	**83.96%**
bicycle	17.97%	26.82%	27.56%	31.11%	37.31%	36.47%
bird	50.78%	56.21%	36.70%	83.32%	80.55%	**84.82%**
boat	45.04%	47.12%	38.74%	65.16%	64.34%	**66.74%**
bottle	41.41%	48.64%	43.73%	70.62%	68.47%	68.22%
bus	69.10%	70.60%	65.79%	84.98%	86.41%	**87.69%**
car	60.73%	60.72%	58.34%	76.56%	81.82%	80.18%
cat	56.51%	59.85%	63.79%	82.68%	85.41%	85.28%
chair	11.66%	20.82%	24.00%	32.31%	32.75%	**33.26%**
cow	52.84%	42.11%	64.85%	69.65%	77.23%	**77.25%**
dining table	19.41%	38.69%	41.34%	54.08%	52.42%	**55.82%**
dog	49.00%	51.54%	55.19%	77.15%	79.01%	**79.85%**
horse	43.90%	43.65%	59.00%	69.76%	80.03%	73.87%
motorbike	52.86%	52.30%	56.91%	68.51%	73.51%	72.62%
person	46.71%	61.65%	49.11%	78.78%	77.23%	78.11%
potted plant	40.56%	37.36%	31.41%	46.11%	55.94%	52.76%
sheep	49.29%	51.80%	53.60%	77.59%	76.05%	76.18%
sofa	26.23%	22.38%	38.60%	45.25%	37.63%	**45.89%**
train	58.32%	56.29%	53.91%	77.19%	79.83%	**80.54%**
tv monitor	48.20%	57.70%	31.79%	55.34%	68.38%	57.44%
total	47.05%	50.10%	48.48%	67.65%	69.89%	**69.95%**

Figure 3: Comparison of segmentation results of the five algorithms and the algorithm selector

2.3 High level description attributes in algorithm selection

In the scope of this project, we decided to test the hypothesis that high level attributes might contribute significantly to the accuracy of algorithm selection. The work in this section would serve as a proof of concept and as a ground for future research directions.

We suggest that a high-level attribute can be described by a certain set of image features. Thus, if there is a way to extract such attributes from an image, it will be possible to use them in the algorithm selection task.

In this experiment, we manually assigned attributes to 50 objects in images. Those objects were chosen such that each algorithm had an equal amount of image objects for which it performs best compared to all other algorithms. The attributes included image exposure, light types, sharpness of an object and its background, the type of background, etc. Then we trained an ANN and tested the accuracy of selection when using only CNN features, only attributes and CNN features and attributes together as an input. Figure 4 demonstrates the results.

CNN features	22%
Attributes	30%
CNN features and attributes	44%

Figure 4: Performance of algorithm selection using the parameters specified

It can be clearly seen that attributes improve the accuracy of selection significantly. Although this experiment was performed on a small amount of data, it does look promising that finding a way to automatically generate such attributes would help to improve the accuracy of selection. Further research is needed in this direction to study the ways of computing such attributes.

3 Conclusion

This work shows how algorithm selection approach can be useful in the field of machine vision and demonstrates a number of things:

1. A higher selection accuracy can be achieved with using an artificial neural network rather than a support vector machine.
2. Features extracted by convolutional neural networks contribute significantly to the accuracy of selection.
3. Algorithm selection approach outperforms the best algorithm in semantic segmentation.
4. It was shown that high-level attributes can significantly contribute to the accuracy of algorithm selection thus providing a base for future research in this direction.

Literature

1. M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. http://www.pascalnetwork.org/challenges/VOC/voc2012/workshop/index.html.
2. G. Passino, I. Patras, and E. Izquierdo. Aspect coherence for graph-based semantic image labelling. IET Computer Vision, 4(3):183 – 194, 2010.
3. J. Carreira, F. Li, and C. Sminchisescu. Object recognition by sequential figure-ground ranking. International Journal of Computer Vision, 98(3):243–262, 2012.
4. Jamie Shotton, John Winn, Carsten Rother, and Antonio Criminisi. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. International Journal of Computer Vision, 81(1):2 – 23, 2009.
5. Lubor Ladicky, Chris Russell, Pushmeet Kohli, and Philip H. S. Torr. Graph cut based inference with co-occurrence statistics. In Proceedings of the 11th European Conference on Computer Vision: Part V, ECCV’10, pages 239–253, Berlin, Heidelberg, 2010. Springer-Verlag.
6. Bharath Hariharan, Pablo Arbel´aez, Ross Girshick, and Jitendra Malik. Simultaneous detection and segmentation. In European Conference on Computer Vision (ECCV), 2014.
7. Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. CVPR, November 2015.
8. Mohammadreza Mostajabi, Payman Yadollahpour, and Gregory Shakhnarovich. Feedforward semantic segmentation with zoom-out features. CoRR, abs/1412.0774, 2014.
9. John R. Rice. The algorithm selection problem. Advances in Computers, 15:65–118, 1976.
10. Martin Lukac, Kamila Abdiyeva, and Michitaka Kameyama. Symbolic segmentation using algorithm selection. CoRR, abs/1505.07934, 2015.
11. Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the 22Nd ACM International Conference on Multimedia, MM’14, pages 675–678. ACM, 2014.