Parameter selection for contour saliency map transformations

Sarah Carneiro*
Felipe Belém†
Anderson Rocha‡

Abstract

In image processing, in order to visualize a segmentation result produced by saliency map approaches, a normalizing function is needed. Accordingly, studies have shown that by using a sigmoid function as a pre-filtering step and as a normalization one, can, in some ways, prevent over-segmentation. However, by applying this function, certain parameters need to be defined in order to achieve the best results. Therefore, this study will, with the assistance of machine learning, define parameters that could be more assertive than the previous selected proposition and generate better segmentation events.

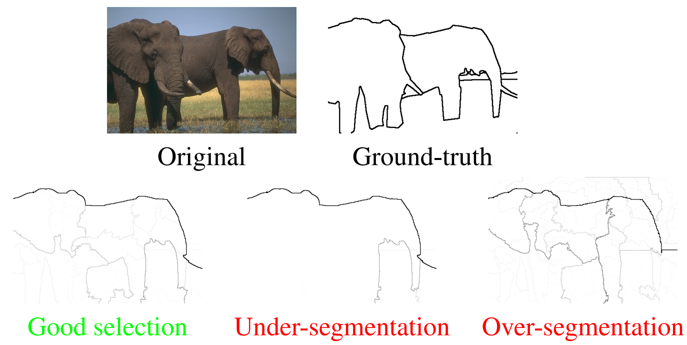

Figure 1. Impacts of the good (or bad) selection of the sigmoidal pre-filtering step, over the saliency map result.

1. Introduction

Image segmentation is the process of trying to contour and distinguish different image objects by grouping similar pixels into regions. Thus, one common method used to perform this task is the watershed algorithm proposed by Beucher and Lantejoul [2]. However, the mentioned method has some drawbacks, especially over-segmentation. This problem consists on the division of single (and relevant) regions, into several irrelevant ones.

To solve this issue, the authors in [6], in order to genuinely define segmentation in terms of a hierarchy of partitions, decided to study the contour saliency map. Belém et. al. [1], in this manner, analyzed the impact of different approaches, in the context of contour saliency map transformations, by using distinct normalizations and a sigmoidal pre-filtering step over the gradient function. In his work, the application of sigmoid functions to the transformation generated better results compared to non-filtered approaches. On the other hand, the sigmoid function used is dependent on a combination of parameters in order to generate good outputs. As it can be seen in Figure 1, the parameter selection impacts significantly in the final result.

Given this context, machine learning has been expanding the possibility of doing multiple testing as well as, in many cases, improving previous state of the art results achieved without the help of a network in various fields of study. In this sense, a technique that can be used, with the help of machine learning considering non discrete data, is regression. Regression can be used as a tool for determining a hypothesis in order to target the expected result by combining features of the given dataset.

The current study proposes the use of regression algorithms to obtain optimal sigmoidal parameters, in the pre-filtering step, of contour saliency map transformations proposed in [1], since its selection in the latter was done empirically. The dataset used during this study was the $BSDS500$, consisting of 261.6 MB of information distributed between 500 $.ppm$ extension images of various scenario situations.

This report will be organized as following. Section 2 will present the regression baseline obtained as well as initial feature selection. In Section 3 we justify the decision to modify the baseline features to run other tests. In Section 4 a regularization procedure is discussed and Section 5 linear regression is substituted by a different approach by using support vector regression (SVR). Section 6 contains the discussion of using a random forest method to cope with the present problem. Section 7 includes the qualitative results obtained by the best configuration obtained during the

*Contact: sarah.alcar@gmail.com
†Contact: felipedkstro@hotmail.com
‡Contact: anderson.rocha@ic.unicamp.br

experimentation. Finally Section 8 we discuss about future studies and the limitations of this study.

2. Baseline

As the authors in Belém et al. [1] concluded, the sigmoid filtering in the gradient has significant impact on the final segmentation result by eliminating minimal valleys which could compromise the construction of the hierarchy, specially if constructed considering the area or the volume as extinction valleys. The counterpart consists on selecting two parameters to determine the curve's shape: (i) steepness; and (ii) middle-point.

Since the definition of good noise filtering is relative to its purpose and is difficult to measure in numbers, the first procedure consisted on establishing, via manual input and measurement, the optimal configurations for each image segmentation. Thus, our approach was based on a semi-supervised solution. Therefore, for a given number of image used as seeds, baseline statistical features such as maximum values, variance, entropy, standard deviation and pixel mean were extracted [7].

The features that were selected try to represent the texture indirectly by non-deterministic properties that define distributions and relationships between the gray levels of the pixels belonging to an image. Therefore, the mean represents the expected value of the gray level distributions present in texture, while variance describes how much the values are scattered around the mean. The entropy will calculate a pixel disorder factor and standard deviation will establish a trust interval to verify the variance between pixels. Lastly the maximum pixel value could be an important border descriptor for our case study.

Therefore, we obtained each image's gradient through the computation of the Euclidean distance between pixels and, since the raw saliency map does not configure as an image, we performed a linear scale normalization on the map (since it maintains the hierarchy relevance), in order to extract its features. Moreover, in order to avoid feature domination, we centered and normalized the features by the L2-norm, which is the default of Scikit-Learn.

As mentioned before, the definition of a good saliency map is rather vague. Therefore, in order to avoid an extremely manual work of selecting the good parameters for every image within the set, we used the K-Means algorithm to group the K closest points in a feature space, with few adaptations. The initial seeds of the algorithm were the images in which we obtained the configurations manually and, since we only desired a simple grouping of samples, we performed only one iteration. After obtaining the clusters, we propagated the seed's values to all the samples within. Although we assume an slight error in the establishment of the labels, the impacts on the final results are also minimal.

After the labels were spread, as a baseline alternative it was decided to use the simplest regression approach, that is linear regression (denoted as LinReg, from now on) [4]. In this sense, both the gradient and the saliency map normalization parameters were obtained separately.

In order to evaluate the performance of every approach, we considered the K-Fold cross validation, considering 10 folds, aiming the minimization of the mean squared error between the prediction value, and the pseudo-label associated to the sample, since it is one of the most used metrics for regression evaluation.

Not surprisingly, the linear regression had a mediocre performance on predicting the values, obtaining an approximate error of 10 units, and 0.1, in the determination of the sigmoid steepness and middle point (as shown in Table 1). However, it is not possible to conclude whether the simplicity of the model, or the features considered, or both, were critical to the baseline performance.

3. Other feature extractions

In order to improve results, tests were conducted by substituting the previous features extracted, that did not consider the relation between pixel neighborhood, to a gray level co-occurrence matrix (GLCM)[7]. The construction of a co-occurrence matrix depends on the transitions of the gray levels between the components of the image, thus, different distances and angles between the pixels end up being included in this matrix.

The information extracted from the GLCM was used as input for the linear regression so its results could be compared to the previous features used during the experiments (this approach is denoted as GLCM, from now on). To build the GLCM the angles used were 0, 30, 45, 60 and 90 degrees with a distance of one pixel. Moreover, the information extracted within each matrix were contrast, dissimilarity, homogeneity, energy, correlation, second angular moment.

Since the feature space was altered, it was needed to spread the labels once more, in order to group the most similar samples in this new domain.

The consideration of the pixels transitions have slightly worsened the results obtained by the baseline. But, as it can be seen in Table 2, the performance was improved on the prediction of the transformation on the saliency map. Thus, it is plausible that the modifications performed improved the predictions overall.

In a different hypothesis, we analyzed if the gradient features, solely, are enough for the prediction of the sigmoid values for saliency map normalization. Interestingly, the performance were significantly improved, as it can be seen in Table 3, indicating that the features extracted from the map may be misleading and, thus, worsening the results.

Fold	Gradient		Saliency map	
	MSE (Steepness)	MSE (Middle-point)	MSE (Steepness)	MSE (Middle-point)
1	168.01	0.0093	261.27	0.02325
2	81.62	0.0056	478.54	0.03312
3	100.16	0.0040	164.90	0.00515
4	96.62	0.0106	480.86	0.02262
5	85.24	0.0045	418.85	0.03761
6	67.95	0.0041	484.42	0.02591
7	102.81	0.0099	343.98	0.01511
8	173.19	0.0070	240.75	0.01953
9	110.62	0.0080	236.28	0.01268
10	65.28	0.0063	801.13	0.05092

Table 1. Results obtained for the prediction of the baseline

Fold	Gradient		Saliency map	
	MSE (Steepness)	MSE (Middle-point)	MSE (Steepness)	MSE (Middle-point)
1	114.96	0.0094	205.01	0.0272
2	84.66	0.0086	523.07	0.0457
3	105.56	0.0129	254.73	0.0246
4	102.56	0.0147	230.63	0.0124
5	109.91	0.0102	379.01	0.0276
6	154.73	0.0170	420.33	0.0185
7	177.95	0.0192	319.14	0.0110
8	113.73	0.0074	411.10	0.0270
9	240.33	0.0120	100.98	0.0067
10	86.34	0.0060	470.22	0.0464

Table 2. Results obtained for the prediction of the linear regression, taking into account texture features

Fold	Saliency map	
	MSE (Steepness)	MSE (Middle-point)
1	138.58	0.01034
2	320.79	0.01944
3	224.76	0.01264
4	270.06	0.0083
5	329.96	0.01436
6	450.06	0.00864
7	243.73	0.00675
8	343.68	0.0110
9	238.28	0.01198
10	368.73	0.02164

Table 3. Results obtained for the prediction of the linear regression, taking into account texture features of the gradient image, only

4. Regularization

Regularization consists on the process of introducing additional information in order to solve overfitting in many models. Therefore, the usage of such will permit an analysis whether our model is overfitting the data and/or avoid it. Through grid search, the best regularization value found was 0.1. However, as it can be seen in Table 4, the final performance was slightly worse than the previous, therefore, we concluded that our model does not overfit the data and discarded this approach.

5. Support vector regression

A different approach tested was by using more complex models compared to a basic linear regression method previously presented. For this step the chosen method was the support vector regression (SVR) [5]. This mechanism has minor differences compared to support vector machines (SVM) used for classification. The idea of an SVR is also based on minimizing the error, maximizing the margin, as well as tolerating part of the error. In the case of the SVR, the tolerance is set as an approximation to the SVM's margin.

As it is possible to see in Table 5, the prediction of the gradient normalization configuration performed fairly better than the previous approach. However, for the saliency map prediction, the performance was decreased. Thus, the strategy considered by the SVR may not be the most adequate for this context.

Fold	Gradient		Saliency map	
	MSE (Steepness)	MSE (Middle-point)	MSE (Steepness)	MSE (Middle-point)
1	92.25	0.00806	136.1067	0.02309
2	74.05	0.00778	353.6160	0.043379
3	124.29	0.0146	234.5681	0.015096
4	101.53	0.014360	179.5397	0.0224
5	94.61	0.0081099	308.46	0.0251069
6	158.45	0.014854	417.478	0.014978
7	142.71	0.021044	205.48836	0.013944
8	109.097	0.008109	364.053	0.024900
9	125.60	0.002291	158.027	0.01396
10	58.36	0.001544	302.364	0.053757

Table 4. Results obtained for the prediction of the regularization method over the GLCM, taking into account texture features

Fold	Gradient		Saliency map	
	MSE (Steepness)	MSE (Middle-point)	MSE (Steepness)	MSE (Middle-point)
1	95.92	0.0114	184.84	0.02314
2	49.46	0.0091	426.88	0.035467
3	101.18	0.0101	219.11	0.02103
4	81.76	0.0115	249.82	0.02325
5	87.46	0.009598	301.36	0.03138
6	139.78	0.008829	367.22	0.01906
7	118.60	0.01399	257.44	0.0.02174
8	86.78	0.008807	352.12	0.02753
9	122.67	0.006915	133.54	0.01981
10	40.78	0.01067	431.06	0.040927

Table 5. Results obtained for the prediction of the SVR, taking into account texture features

6. Random Forest

The expectation of changing once more the regression algorithm is that the results obtained by previous experimentation can be improved. The random forest (denoted as RF, from now on) approach is capable of dealing with both classification and regression tasks [3].

The main concept of this method is to create a forest with a certain number of decision trees. Forests with a greater quantity of trees are, theoretically, expected to generate more robust results, thus, improving accuracy. The forest can be built based on various metrics, one of them being the information gain. Additionally, in the case of regression, given an input each tree will compute an output and the forest computes the final result based on the average of the outputs. An advantage of the random forest is that it can handle large datasets, missing values, can be executed in parallel and can provide feature importance. However, it is important to consider that it is a model that we have little control over it and does not show promising results when dealing with redundant features.

The results obtained by the Random Forest regressor are shown in Table 6. It is possible to see that this model has obtained the best results overall, specially in predicting the values for the saliency map normalization. Thus, it is plausible to infer that a single regressor can predict the optimal parameters for both normalizations (gradient and saliency map) with some confidence.

7. Qualitative Results

In this section qualitative results generated by the previous discussed experiments can be seen.

As it can be seen in Figure 2, each predictor had a different outcome for the given image. For the configuration proposed by Belém et. al., although the bear was well delineated and there was a significant noise suppression in homogeneous zones, some edges were also suppressed in consequence of the parameters considered. The results of LinReg and SVR were similar: it has enhanced the edges in the gradient, but the selection of the normalization parameters impacted severely the segmentation result. Interestingly, a simple linear regression over the texture-based features had a similar result, but it has supressed many relevant edges. The RF approach presented the best result overall by assigning the adequate relevance to the edges, while suppressing the undesired noises

Fold	Gradient		Saliency map	
	MSE (Steepness)	MSE (Middle-point)	MSE (Steepness)	MSE (Middle-point)
1	93.15	0.004847	3.13	8.1699e-05
2	16.65	0.0002418	2.85	6.65e-06
3	30.25	0.00082675	10.7	3.6159e-05
4	25.65	0.0021366	1.1125	2.342e-05
5	42.25	0.00101445	53.675	0.000276
6	66.6	0.00276969	23.1705	0.000143214
7	115.9	0.0060482	27.0165	0.000198562
8	49.5	0.0001378	63.45	0.00254373
9	91.35	0.0006943	28.075	0.001326926
10	62.15	0.00025695	18.644	0.00240749

Table 6. Results obtained for the prediction of the random forest regressor, taking into account texture features

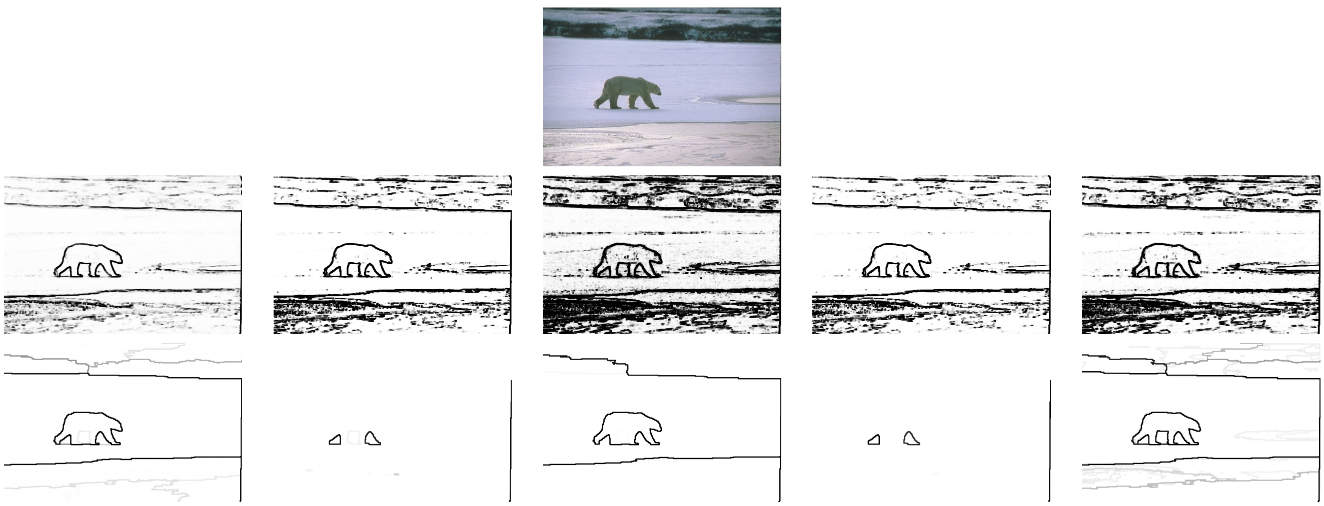

Figure 2. Performance of the algorithm considering each predictors configuration. The image in the first row, is the original image. In the second row consists on the filtered gradient image, and the third, the respective normalized saliency map.

8. Conclusion

During our study, the main objective was to optimize the sigmoid parameters for a more ideal image segmentation (thus, maximizing the reduction of noises). We expected to obtain a similar pattern, but with better results compared to previous empirically tested results. In order to establish an approximation of the best configuration for each image, a semi-supervised approach was considered, such that the regression was made possible. Experimental results show that an optimal parameter can be estimated by a Random Forest model regarding texture-based features as input samples, overcoming other approaches (like SVR), and obtaining a better performance than the result presented by Belém et al. [1].

For future endeavors, it is intended to extend this work for different texture-based features, alongside measuring the performance of the predictor considering different and more efficient gradient functions.

References

[1] F. d. C. Belém, J. Cousty, Z. K. do Patrocınio Jr, and S. J. F. Guimaraes. Impacts of contour saliency map transformations. In *Graphics, Patterns and Images (SIBGRAPI), 2017 30th SIBGRAPI Conference on*. IEEE, 2017. 1, 2, 5

[2] S. Beucher. Use of watersheds in contour detection. In *Proceedings of the International Workshop on Image Processing*. CCETT, 1979. 1

[3] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001. 4

[4] M. B. Christopher. *PATTERN RECOGNITION AND MACHINE LEARNING*. Springer-Verlag New York, 2016. 2

[5] H. Drucker, C. J. Burges, L. Kaufman, A. J. Smola, and V. Vapnik. Support vector regression machines. In *Ad-*

vances in neural information processing systems, pages 155–161, 1997. 3

[6] L. Najman and M. Schmitt. Geodesic saliency of watershed contours and hierarchical segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 18(12):1163–1173, 1996. 1

[7] H. Pedrini and W. R. Schwartz. *Análise de Imagens Digitais*. Thomson, 2007. 2