Textonboost based on Differential Evolution

Keiko Ono
Ryukoku University
Kyoto, Japan
kono@rins.ryukoku.ac.jp

Daisuke Tawara
Ryukoku University
Kyoto, Japan
tawara@rins.ryukoku.ac.jp

Yoshiko Hanada
Kansai University
Osaka, Japan
hanada@kansai-u.ac.jp

ABSTRACT

Semantic segmentation on a pixel basis is necessary for the semantic understanding of an image. Although the use of CNN is mainstream in the case where there are sufficient test images, in this research we aim to develop a method that is robust in an environment with few test images. Specifically, we improve Textonboost by using a differential evolution method.

CCS CONCEPTS

• **Mathematics of computing** → *Evolutionary algorithms*;

KEYWORDS

Image Segmentation, Textonboost, Differential Evolution

ACM Reference Format:
Keiko Ono, Daisuke Tawara, and Yoshiko Hanada. 2019. Textonboost based on Differential Evolution. In *Genetic and Evolutionary Computation Conference Companion (GECCO '19 Companion), July 13–17, 2019, Prague, Czech Republic.* ACM, New York, NY, USA, 2 pages.
https://doi.org/10.1145/3319619.3321896

1 INTRODUCTION

Extraction of attention regions in images is an important research area in the field of image processing. Within this area, semantic segmentation is a region extraction problem that requires precise accuracy for the purpose of semantic understanding of pixels. At present, Convolutional Neural Networks (CNN) typified by SegNet are often used for semantic segmentation, but since it is necessary to learn specific weighting parameters between multiple layers, it is common to use several hundred images for SegNet learning, and high extraction performance cannot be expected when many learning images are not available. In such a case, image processing methods combining clustering methods such as Mean-Shift and K-means are often used, and Textonboost[4] has high performance. Textonboost is used not for standalone processing, but also for CNN post-processing[1].

Textonboost estimates the class for each pixel by boosting, but differs from other methods by greatly reducing calculation costs, by not directly compressing high order image features, but by dimensionally compressing them into feature quantities of a two-dimensional map called a Texton Map. However, the importance of each feature quantity is not taken into account. In this research, we aim to improve performance by considering importance for each feature quantity. The feature quantity is a response to the image filter, and is represented by a real value. There are various high-dimensional real-valued optimization methods, but we consider the application of Differential Evolution(DE), which can quickly find a quasi-optimal solution when there is dependence between design variables and there are multiple local optima[3].

GECCO '19 Companion, July 13–17, 2019, Prague, Czech Republic

ACM ISBN 978-1-4503-6748-6/19/07…$15.00
https://doi.org/10.1145/3319619.3321896

2 TEXTONBOOST

Consider the semantic segmentation of each pixel p_j when the class $c(p_j)$ of the correct data is given to p_j of the image set I. The algorithm is shown below.

2.1 Algorithm

O-0: Initialize Divide I into learning and test data $I^{learn}$and I^{test}, respectively. Convert the I^{learn} RGB colormetric image $\boldsymbol{m}_h;\ h = 1, \ldots H$ to the LAD colormetric system $\boldsymbol{l}_h;\ h = 1, \ldots, H$.

O-1: Execute Filterbank Apply the filter $\boldsymbol{b}_k;\ k = 1, \ldots, K$ to $\boldsymbol{l}_h$ and obtain filler response L, where
$L = \{\{\boldsymbol{l}_1(\boldsymbol{b}_1), \ldots, \boldsymbol{l}_1(\boldsymbol{b}_K)\}, \ldots, \{\boldsymbol{l}_H(\boldsymbol{b}_1), \ldots, \boldsymbol{l}_H(\boldsymbol{b}_K)\}\}$.

O-2: Execute K-Means L is divided into two: L_1 and L_2. Execute K-means on L_1 to extract centroids $c_1, \ldots, c_\alpha$.

O-3: Generate Textonmap Execute NN on L_2 and assign it to the nearest centroid. As a result Textonmap $\boldsymbol{tm}_h;\ h = 1, \ldots H$ is generated.

O-4: Calculate Local Features Generate filter windows w_1, $\ldots, w_W$ in order to extract local features. In order to extract M features, the following processing is repeated M times.
(1)$m \leftarrow 0$
(2)Select an arbitrary filter window and Textonmap.
(3)Calculate the ratio (local feature quantity) of class $1, \ldots, \alpha$ in the window for $\boldsymbol{tm}_h$[1].
(4)$m \leftarrow m + 1$

O-5: Execute Boosting Execute boosting based on the M features of α dimensions extracted in Step O-4.

O-6: Evaluate Apply the rule with the highest accuracy to the I^{test}.

3 PROPOSED METHOD

Textonboost compresses high-dimensional features of an image into Textonmap, which is a two-dimensional feature quantity, by K-means and NN. Textonmap holds the label of each pixel, and learns the class with respect to the pixel in a test image by boosting of the weak classifier, using the relationship between the abundance ratio for the class of the local area and the class at the center

[1]Since window size differs individually, it is normalized by the window size.

of the area as learning data. The high dimensional feature is the response quantity of each filter, but all the filters are treated with the same weighting. However, since the response differs by the filter structure, performance improvement can be expected by optimizing the filter effective for semantic segmentation. In this research, we propose a method to learn the importance of filter which was not taken into account in Textonboost. Since the weighting is a real number and becomes the dimension of the filter number, we propose application of differential evolution (DE), which is an easy-to-implement method with high performance, in the optimization of high-dimensional real values.

3.1 Algorithm

Step O-1of Textonboost is changed as follows and the importance of the response of each filter is learned by AdaptiveDE.

- **P-0** Initialize the initial weight $\boldsymbol{x}^i = \{x_1^i, \ldots, x_K^i\}$ with a uniform random number $U(0, 1)$ for each filter to generate the initial population X.
- **P-1** Execute AdaptiveDE. The objective function $f(\boldsymbol{x})$ is shown in Section 3.2.
- **P-2** Find the weight $\boldsymbol{x}^*$ of the individual with the lowest $f(x)$.
- **P-3** Update the filter response.

$$L \leftarrow \{\{x_1^* \cdot \boldsymbol{l}_1(\boldsymbol{b}_1), \ldots, x_K^* \cdot \boldsymbol{l}_1(\boldsymbol{b}_K)\}, \ldots, \{x_1^* \boldsymbol{l}_H(\boldsymbol{b}_1), \ldots, x_K^* \cdot \boldsymbol{l}_H(\boldsymbol{b}_K)\}\}$$

3.2 Objective Function

Textonboost uses K-means of Unsupervised Learning, but since the target problem has a correct label for each pixel, clustering performance can be verified. However, since there is no relationship between the result of K-means and the correct data, it is evaluated as follows.

Let the pixels of the learning image $\boldsymbol{I}_h$ be p_j; $j = 1, \ldots, J$, and generate a correct set $C^* = C_1^* \cup \cdots \cup C_\alpha^*$ from the class label of the correct data given to each pixel p_j. Here, $p_j \in C_t^*$ if $c(p_j) = t$ $(t = 1, \ldots, \alpha)$ and J = number of pixels × H. Likewise, a set of pixels $C = C_1 \cup \cdots \cup C_\alpha$ allocated to each cluster of K-means is generated. Compare pixels belonging to an arbitrary correct class C_t^* with class set C, and let the aim be to find $\boldsymbol{x}_i$ where the difference from class C_{t^*} with the most common sections is small. The objective number is defined as follows.

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}_i \in X} f(\boldsymbol{x}_i),$$

$$f(\boldsymbol{x}_i | C^*, C) = \sum_t C_t^* \setminus C_{t^*},$$

$$C_{t^*} = \arg\max_{C_t \in C} C_t^* \cap C_t.$$

4 PERFORMANCE EVALUATION

In this paper, 3 types of two-dimensional Gaussian filters with the standard deviation $\sigma = \{\sqrt{2}, 2\sqrt{2}, 4\sqrt{2}\}$ out of the 48 filters of [2] and 12 types of one-dimensional Gaussian filter with rotation angles $\{\pi/3, \pi/2, \pi\}$ are adopted. Filter size can be arbitrarily set, but it is set to 7*7 in the experiment. Figure 1 shows the data set. It was divided into two: learning and testing. For the DE parameter, the number of individuals is 10 and the generation number is 10. In

The results of Boosting are shown in Table 1. From this result, it was found that the proposed method that divides each dimension

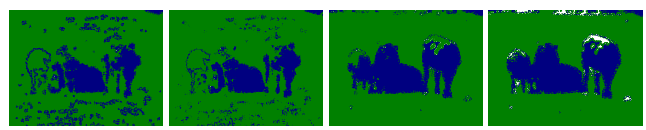
order to confirm the performance of the proposed method, performance is evaluated by four methods: the Textonboost of the original, the method of adding DE to the original, the Proposed-DE using the filter response of the Lab without DE for each component, and the DE using both.

Figure 1: Test Data.

Table 1: Accuracy (Boosting).

Methods	original	original+DE	proposed-DE	proposed
Fitness	0.854	0.849	0.861	0.874

of Lab and learned the weighting using DE had the highest accuracy. Also, Figure 2 shows the test results. From this result also, it was found that the proposal has the highest segmentation result.



(a) original (b) original+DE (c) proposed-DE (d) proposed

Figure 2: Estimated regions.

5 CONCLUSIONS

In this research, we aimed to develop a method robust to environments with few test images, and proposed a method combining a differential evolution method and Textonboost. The results showed that the proposed method demonstrates the highest performance.

ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Number JP19H02032 and a research project of 2019 Joint Research Center for Science and Technology Fund of Ryukoku University.

REFERENCES

[1] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. 2018. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence* 40, 4 (2018), 834–848.

[2] Thomas Leung and Jitendra Malik. 2001. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International journal of computer vision* 43, 1 (2001), 29–44.

[3] A Kai Qin, Vicky Ling Huang, and Ponnuthurai N Suganthan. 2009. Differential evolution algorithm with strategy adaptation for global numerical optimization. *IEEE transactions on Evolutionary Computation* 13, 2 (2009), 398–417.

[4] Jamie Shotton, John Winn, Carsten Rother, and Antonio Criminisi. 2006. Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. (2006), 1–15.