Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn



Interactive image segmentation with a regression based ensemble learning paradigm*

Jin ZHANG[†1], Zhao-hui TANG[†‡1], Wei-hua GUI[1], Qing CHEN[2], Jin-ping LIU[3]

([1]*School of Information Science and Engineering, Central South University, Changsha 410083, China*)

([2]*College of Computer and Communication, Hunan University of Technology, Zhuzhou 412007, China*)

([3]*College of Mathematics and Computer Science, Hunan Normal University, Changsha 410083, China*)

[†]E-mail: zhang_jin@csu.edu.cn; zhtang@csu.edu.cn

Received July 18, 2016; Revision accepted Mar. 14, 2017; Crosschecked July 11, 2017

Abstract: To achieve fine segmentation of complex natural images, people often resort to an interactive segmentation paradigm, since fully automatic methods often fail to obtain a result consistent with the ground truth. However, when the foreground and background share some similar areas in color, the fine segmentation result of conventional interactive methods usually relies on the increase of manual labels. This paper presents a novel interactive image segmentation method via a regression-based ensemble model with semi-supervised learning. The task is formulated as a non-linear problem integrating two complementary spline regressors and strengthening the robustness of each regressor via semi-supervised learning. First, two spline regressors with a complementary nature are constructed based on multivariate adaptive regression splines (MARS) and smooth thin plate spline regression (TPSR). Then, a regressor boosting method based on a clustering hypothesis and semi-supervised learning is proposed to assist the training of MARS and TPSR by using the region segmentation information contained in unlabeled pixels. Next, a support vector regression (SVR) based decision fusion model is adopted to integrate the results of MARS and TPSR. Finally, the GraphCut is introduced and combined with the SVR ensemble results to achieve image segmentation. Extensive experimental results on benchmark datasets of BSDS500 and Pascal VOC have demonstrated the effectiveness of our method, and the comparison with experiment results has validated that the proposed method is comparable with the state-of-the-art methods for interactive natural image segmentation.

Key words: Interactive image segmentation; Multivariate adaptive regression splines (MARS); Ensemble learning; Thin-plate spline regression (TPSR); Semi-supervised learning; Support vector regression (SVR)

http://dx.doi.org/10.1631/FITEE.1601401 **CLC number:** TP391.4

1 Introduction

Image segmentation is one of the most fundamental problems in low-level visual processing. It is the basis of follow-up target extraction, image analysis, recognition, and understanding. A large number of automatic image segmentation methods have been proposed. However, it is a difficult task to achieve automatic segmentation of foreground objects from the complex background. Currently, interactive segmentation methods (Yang *et al*., 2010; Peng *et al*., 2013; Jung *et al*., 2014; Jian and Jung, 2016; Wang *et al*., 2016) incorporating simple manual interaction (manually adding some prior knowledge about region segmentation) have been actively studied. Interactive image segmentation aims at compensating for the low efficiency of manual segmentation and the low precision of fully automatic segmentation.

The interactive image segmentation model can be trained after separately and manually pointing out a part of the foreground pixels $\{L_{\boldsymbol{O}_\mathrm{F}}|\boldsymbol{O}_\mathrm{F}=[x_1^{\boldsymbol{O}_\mathrm{F}}, x_2^{\boldsymbol{O}_\mathrm{F}}, \ldots, x_{N_{\boldsymbol{O}_\mathrm{F}}}^{\boldsymbol{O}_\mathrm{F}}]\}$ and a part of the background pixels $\{L_{\boldsymbol{O}_\mathrm{B}}|\boldsymbol{O}_\mathrm{B}=[x_1^{\boldsymbol{O}_\mathrm{B}}, x_2^{\boldsymbol{O}_\mathrm{B}}, \ldots, x_{N_{\boldsymbol{O}_\mathrm{B}}}^{\boldsymbol{O}_\mathrm{B}}]\}$. Then a reasonable area

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 61071176, 61171192, and 61272337) and the Doctoral Fund of the Ministry of Education of China (No. 20130162110013)

ORCID: Jin ZHANG, http://orcid.org/0000-0001-7574-2808

© Zhejiang University and Springer-Verlag Berlin Heidelberg 2017

label can be allocated to those unlabeled pixels $U=\{x_1, x_2, \ldots, x_M\}$ automatically by a segmentation model (where $\boldsymbol{O}_\mathrm{F}$ and $\boldsymbol{O}_\mathrm{B}$ represent foreground and background respectively, $N_{\boldsymbol{O}_\mathrm{F}}$ and $N_{\boldsymbol{O}_\mathrm{B}}$ represent the labeled pixels of foreground and background, and $x_i^{\boldsymbol{O}_\mathrm{F}}$ and $x_i^{\boldsymbol{O}_\mathrm{B}}$ represent the foreground pixel and background pixel). Up to the present, a variety of interactive image segmentation methods have been proposed. GraphCut (GC) (Zhou *et al*., 2014; Pauchard *et al*., 2016) is one of the most classical methods and a variety of improved versions, such as GrabCut (Rother *et al*., 2004), lazy snapping (Li *et al*., 2004), hierarchical graph cuts (Zhang *et al*., 2010; Ge *et al*., 2015), and OneCut (Tang *et al*., 2013), have been proposed. However, despite the success of these existing approaches, they all have a noticeable level of limitation for those images whose foreground and background share similar color distributions such as in cluttered or camouflaged images (Yang *et al*., 2010; Wu *et al*., 2014). Region merging (Ning *et al*., 2010), geodesic methods (Gulshan *et al*., 2010), and active contours (Liu and Yu, 2012; Nguyen *et al*., 2012) are also important interactive segmentation algorithms and have received much attention by many researchers.

From the perspective of machine learning, the image segmentation problem is usually viewed as a problem of pattern classification and label regression (Wu *et al*., 2014; Ding *et al*., 2015). There have been a variety of relevant pattern recognition methods, such as support vector machine (SVM) (Wang *et al*., 2011a; 2011b), spline regression (SR) (Xiang *et al*., 2009; 2010), and random walk (RW) (Qin *et al*., 2014; Zhou and Garcia, 2016). Because of the lack of emphasis on the structure information of pixel features, SVM methods can hardly complete fine segmentation of those images with similar areas existing in both the foreground and background (Xiang *et al*., 2009). The segmentation accuracy on the complex images of SR (Xiang *et al*., 2009) has greatly improved compared with the SVM method, GC (Kolmogorov and Zabih, 2004), and RW (Qin *et al*., 2014; Zhou and Garcia, 2016). However, sometimes the fine segmentation can rely on adding a large number of manual labels. Unfortunately, detailed manual labeling will lead to a huge expenditure of resources and bring great inconvenience to users. In addition, it has been reported that the less the user interaction, the more robust the segmentation model (Ning *et al*., 2010).

After a detailed analysis of the methods proposed in the literature (Xiang *et al*., 2009; 2010; Jobst *et al*., 2016), we have found that there is a key problem in smooth thin plate spline regression (TPSR). It ignores the local interaction among feature factors that contain some more plentiful classification information (in this paper, any one-dimensional value in the feature vector is named a 'feature factor'). This will lead to a poor segmentation when manually labeled pixels are relatively scarce compared to the image feature patterns contained in the image. Multivariate adaptive regression splines (MARS) (Friedman, 1991) is a classical regression method and has been applied to a variety of prediction problems (Adamowski *et al*., 2012; Menon *et al*., 2014; Zhang and Goh, 2016). MARS is a function of every single factor and interactions between multiple feature factors. Theoretically, it can compensate for the deficiency of TPSR. Therefore, an ensemble method based on TPSR and MARS is proposed to achieve much more robust interactive image segmentation results.

In interactive image segmentation, there are many more unlabeled pixels than labeled pixels. If these unlabeled pixels can be fully used to assist the training of the segmentation model, the segmentation performance will be greatly improved. Semi-supervised learning can be roughly divided into three categories:

1. Semi-supervised learning method based on the generative model. The probability of the category of the unlabeled pixels is taken as a default parameter. Then the expectation maximization (EM) algorithm is adopted to estimate the parameters of this model. The commonly used models include the Gaussian mixture model (Shahshahani and Landgrebe, 1994) and mixture of experts (Lee and Cho, 2014). Although this method is simple and intuitive, its accuracy relies too much on the choice of the model structure.

2. Semi-supervised learning method based on a data graph (Xiang *et al*., 2010). First, the labeled pixels and unlabeled pixels are used to construct the data graph structure, and then the label will be propagated to the unlabeled pixels from labeled pixels based on the adjacency relations in the graph. The performance of this method also depends too much on the construction method of the data graph.

3. Co-training (Zhou and Li, 2007). Co-training is aimed to train two classifiers separately on two different views, i.e., two independent sets of attributes, and to use the predictions of each classifier on unlabeled examples to augment the training set of the other. Since Blum and Mitchell (1998) presented the original co-training algorithm, many variations have been proposed (Balcan *et al*., 2004; Zhou and Li, 2005; Zhang *et al*., 2014). These methods tend to focus on the research on the classifier itself, whereas the application to interactive image segmentation has been rarely studied.

There is a strongly complementary nature in the ensemble model of MARS and TPSR, and these two individual models can be trained independently. Thus, the necessary condition that there be two independent views of sufficiency and redundancy in co-training is satisfied. Therefore, a new semi-supervised learning paradigm method named 'regressor boosting' is proposed to further assist the training of the ensemble model using unlabeled pixels.

2 Interactive image segmentation via regression based ensemble learning

This paper presents a novel interactive image segmentation method via regression based ensemble learning. First, two spline regressors MARS and TPSR with the complementary characteristics are trained independently based on manually labeled pixels. Then, to make full use of classification information contained in unlabeled pixels to refine the two individual models MARS and TPSR, a regressor boosting method based on the theory of boosting strategy (Galar *et al*., 2012) is proposed. Next, a decision fusion strategy based on support vector regression (SVR) (Opitz and Maclin, 1999; Lazaridis *et al*., 2011; Fu *et al*., 2014) is introduced to integrate the regression results of MARS and TPSR. It lacks a spatial affinity constraint in per-pixel regression based image segmentation, and this may make it prone to produce noisy and unaligned segmentation in some situations. In the final step, GC (Kolmogorov and Zabih, 2004) is introduced to augment the segmentation result by combining with a boundary constraint. This is achieved by setting the outputs of SVR as a regional penalty and using an ad-hoc function to construct a boundary penalty. For simplicity of description, we use Reg_EL to represent the regressor ensemble segmentation method without GC augmentation and Reg_EL_GC to represent the regressor ensemble segmentation method augmented by GC. The performance of the proposed Reg_EL_GC has been validated on several classic image datasets, and the effectiveness and superiority of this model have been analyzed and compared with some state-of-the-art methods in interactive image segmentation.

2.1 Description of the image segmentation problem

Assuming that the training set contains $N_{\boldsymbol{O}_\mathrm{F}}+N_{\boldsymbol{O}_\mathrm{B}}$ manually labeled pixels, which are obtained by drawing green and red lines on the foreground and background separately (the unmarked regions are viewed as unlabeled pixels and need to be segmented). For convenience, we represent the pixel features at the labeled points as $\{x_{\boldsymbol{O}_\mathrm{F}}{}^1, x_{\boldsymbol{O}_\mathrm{F}}{}^2, \ldots, x_{\boldsymbol{O}_\mathrm{F}}{}^{N_{\boldsymbol{O}_\mathrm{F}}}\}\cup\{x_{\boldsymbol{O}_\mathrm{B}}{}^1, x_{\boldsymbol{O}_\mathrm{B}}{}^2, \ldots, x_{\boldsymbol{O}_\mathrm{B}}{}^{N_{\boldsymbol{O}_\mathrm{B}}}\}$. The region segmentation label response corresponding to the observation pixel x_i is y_i, $y_i\in\{\boldsymbol{O}_\mathrm{F}, \boldsymbol{O}_\mathrm{B}\}$, and we have

$$y=\begin{cases}+1, & y\in\boldsymbol{O}_\mathrm{F},\\ -1, & y\in\boldsymbol{O}_\mathrm{B}.\end{cases} \tag{1}$$

The task of interactive image segmentation is to establish a regression model between the label response and input pixel feature vectors, namely

$$y_i=h(x_i)+\varepsilon, \tag{2}$$

where ε represents the model residual error. Therefore, the key point of this study is to make full use of the manually labeled pixels and a large number of unlabeled pixels to construct an effective approximation function $\hat{h}$ of $h(x)$. Under the ideal condition, $\hat{h}(x_i^{\boldsymbol{O}_F})=1$, $\hat{h}(x_i^{\boldsymbol{O}_B})=-1$. However, the estimated label value actually meets the distribution (Figs. 1b, 1d, and 1e). To accomplish the above classification task, take 0 as the threshold. The pixel that satisfies $\hat{h}(x_i)>0$ will be judged as the foreground pixel, while the pixel that satisfies $\hat{h}(x_i)<0$ will be judged as the background pixel.

To obtain the optimal estimate of h is actually to obtain the similar approximation function of Fig. 1e; namely, the distribution of the predicted value obtained

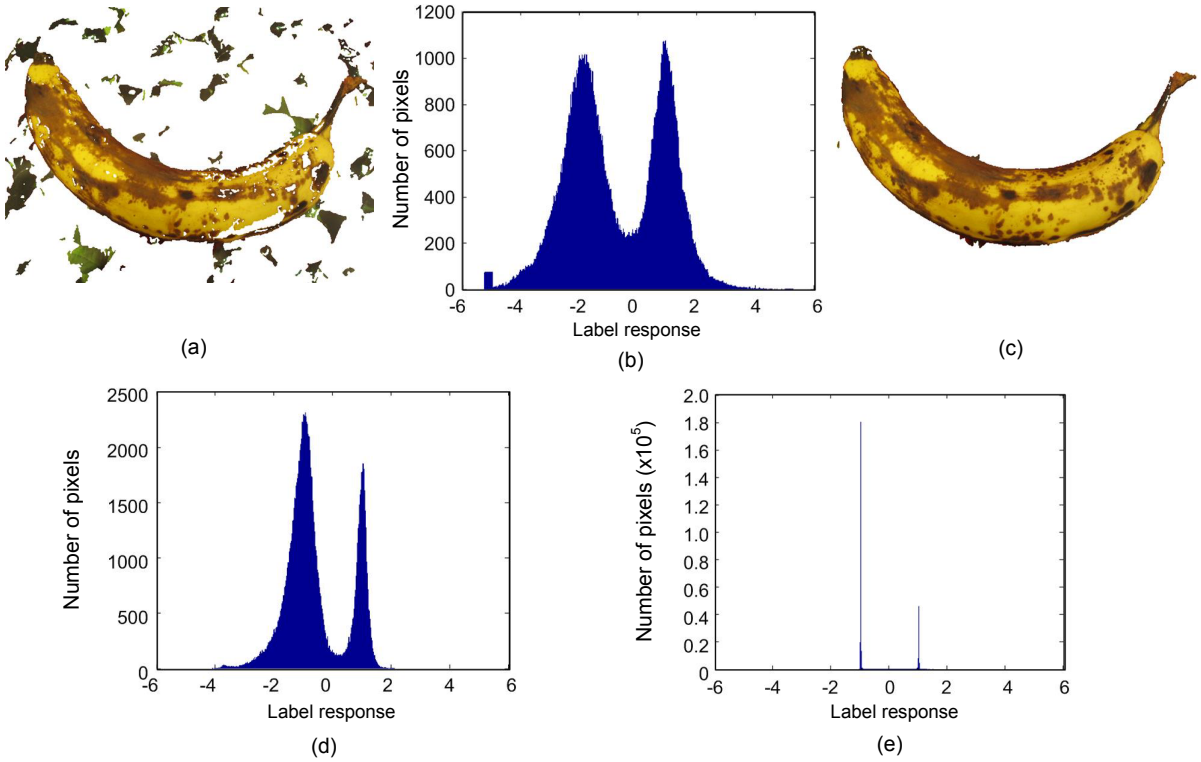


Fig. 1 Histogram of label response: (a) segmentation result of TPSR, with k=7, M'=0; (b) label response histogram of (a); (c) segmentation result of TPSR, with k=65, M'=8000; (d) label response histogram of (c) by TPSR; (e) label response histogram of (d) by Reg_EL

by $\hat{y}$ should have bimodal characteristics, where the peak is narrow, tall, and straight, the span of the peak base is narrow, and two peak bases are non-sticky.

2.2 Ensemble of two complementary regressors

The general form of ensemble regressor $H(x)$ is set as $H(x)=\omega_1 h_1(x)+\omega_2 h_2(x)$, where ω_i $(0<\omega_i<1)$ is the weight of regressor h_i. $h_1(x)$ and $h_2(x)$ are built from two independent and complementary views. Specifically, regressor $h_1(x)$ is constructed based on MARS (Friedman, 1991), which is sensitive to noise but makes full use of the single feature factor contribution and the interactions among multiple feature factors. Regressor $h_2(x)$ is constructed based on the smooth TPSR (Xiang *et al*., 2009), which obtains the classification information from the view of structure information in pixel features.

In prior experiments, we found that if the values of ω_1 and ω_2 were simply and manually set, there would be some wrongly segmented small areas by $H(x)$, though they were correctly segmented by one of regressors $h_1(x)$ and $h_2(x)$. This is because the correct label response of one regressor is a small value, while the wrong label response of the other regressor is a larger one with opposite sign. In this condition, it is hard to obtain a reasonable ensemble value by taking fixed weights ω_1 and ω_2. Fig. 2 further indicates that the label response relationship between $h_1(x)$ and $h_2(x)$ is non-linear. Therefore, in this study we introduce an SVR-based fusion model to ensemble $h_1(x)$ and $h_2(x)$.

2.2.1 Multivariate adaptive regression spline

The MARS method is a kind of non-linear and nonparametric local regression method (Friedman, 1991; Menon *et al*., 2014), which simulates the complex non-linear relationship through the spline function. The general form of MARS is as follows:

$$h_1(x)=\lambda_0+\sum_{m=1}^{M}\lambda_m\cdot\prod_{k=1}^{K_m}\left[s_{km}\left(x_{\upsilon(k,m)}-t_{km}\right)\right]_+, \quad (3)$$

where M represents the spline base number, $\lambda=\{\lambda_0, \lambda_1, \ldots, \lambda_M\}$ are the output weights, K_m is the segmentation times of the mth spline, and t_{km} is the spline node or threshold for segmentation. $\upsilon(k, m)\in\{1, 2, \ldots, p\}$ indicates which one of the p feature factors enters

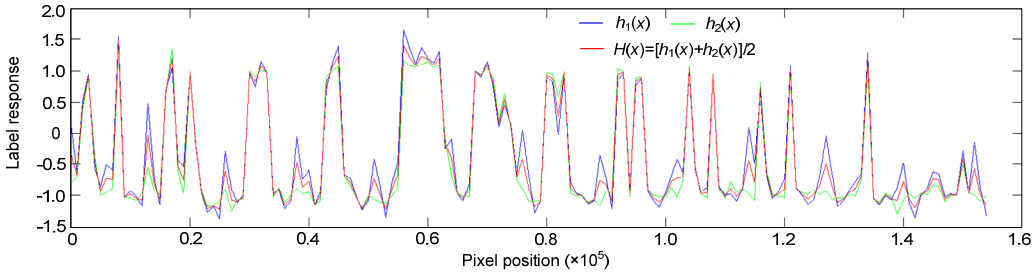


Fig. 2 Label responses of a natural image after flattening it into a column vector

into the kth subsection of the mth spline. $s_{km}\in\{+1, -1\}$ determines the direction of spline segmentation. $[s_{km}(x_{\upsilon(k,m)}-t_{km})]_+$ is a half polynomial, namely

$$[s_{km}(x_{\upsilon(k,m)}-t_{km})]_+=\begin{cases}x_{\upsilon(k,m)}-t_{km}, & \text{if } x_{\upsilon(k,m)}>t_{km},\\ 0, & \text{otherwise.}\end{cases} \quad (4)$$

The MARS formula $h_1(x)$ is broken down into the following intuitional form, which is easily analyzed by ANOVA (analysis of variance) (Friedman, 1991):

$$\hat{h}_1(x)=\lambda_0+\sum\nolimits_{K_m=1}h_{1_i}(x_i)+\sum\nolimits_{K_m=2}h_{1_{ij}}(x_i,y_j)+\sum\nolimits_{K_m=3}h_{1_{ijk}}(x_i,y_j,x_k)+\cdots, \quad (5)$$

where K_m represents the number of feature factors. The first accumulation item is the sum of all basis function results containing only a single feature factor effect, the second accumulation item is the sum of all basis function results containing only interaction between two feature factors, and so on, such that the mth accumulation item is the sum of all basis function results containing only interactions among m feature factors.

Eq. (5) shows that the MARS regression model not only uses the contribution of a single factor, but also takes full advantage of the interactions among multiple factors. Furthermore, it explores the complex information structure hidden in depth in the multi-dimensional feature. Consequently, when training samples are lacking, the MARS method is superior to similar classification algorithms as it makes full use of the feature information of the manually labeled sample pixels from various views.

2.2.2 Smooth thin plate spline regression

TPSR (Xiang *et al*., 2009; 2010) is constructed from the Sobolev space taking both interpolation accuracy and spline smoothness into comprehensive consideration, which can be achieved by minimizing function $J(h_2)$:

$$J(h_2)=\sum_{i=1}^{N_{\boldsymbol{o}_\mathrm{F}}}\left[y_i-h_2(x_i^{\boldsymbol{o}_\mathrm{F}})\right]^2+\sum_{i=1}^{N_{\boldsymbol{o}_\mathrm{B}}}\left[y_i-h_2(x_i^{\boldsymbol{o}_\mathrm{B}})\right]^2+\lambda S(h_2), \quad (6)$$

where $S(h_2)$ is the smoothness penalty function of h_2:

$$S(h_2)=\int_{\Re^d}\sum_{i=1}^{d}\sum_{j=1,j\neq i}^{d}(h_{2_{x_ix_i}}^2+h_{2_{x_ix_j}}^2+h_{2_{x_jx_j}}^2)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_d. \quad (7)$$

Studies indicate that the solution which satisfies Eq. (6) under constraint condition (7) is the following Green formula:

$$\varphi(r)=\begin{cases}r^{4-d}\ln r, & d=2 \text{ or } 4,\\ r^{4-d}, & \text{otherwise,}\end{cases} \quad (8)$$

where $r=\|x-x_i\|$. Thus, TPSR classifier $h_2(x)$ can be obtained through the following formula:

$$h_2(x)=\beta_0+\sum_{i=1}^{d}\beta_i x_i+\sum_{j=1}^{N_{\boldsymbol{o}_\mathrm{F}}}\alpha_j^{\boldsymbol{o}_\mathrm{F}}\phi_j^{\boldsymbol{o}_\mathrm{F}}(x)+\sum_{j=1}^{N_{\boldsymbol{o}_\mathrm{B}}}\alpha_j^{\boldsymbol{o}_\mathrm{B}}\phi_j^{\boldsymbol{o}_\mathrm{B}}(x), \quad (9)$$

where $\varphi_j^{\boldsymbol{o}_\mathrm{F}}=\|x-x_j^{\boldsymbol{o}_\mathrm{F}}\|^2\log\|x-x_j^{\boldsymbol{o}_\mathrm{F}}\|$, and $\varphi_j^{\boldsymbol{o}_\mathrm{B}}=\|x-x_j^{\boldsymbol{o}_\mathrm{B}}\|^2\cdot\log\|x-x_j^{\boldsymbol{o}_\mathrm{B}}\|$.

The Green solution $\phi(r)$ in the TPSR method is associated with the labeled training pixels. Since $r=\|x-x_i\|$ is the spatial distance between x and x_i, we can conclude that the TPSR method is constructed based on the space distribution of the sample pixels. Though it is different from the MARS method, there

is no emphasis on the classification information contained in the singular feature factor and the interactions among several feature factors (the number of feature factors is greater than two but less than d). TPSR focuses on the structure information of pixel features. Therefore, it can make up for the shortages of the MARS method in feature structure information and this is the reason why we declare that there is a complementarity in nature in the ensemble models of MARS and TPSR.

2.2.3 SVR based regressor ensemble

The two individual models of MARS regressor h_1 and TPSR regressor h_2 are trained in advance. The SVM based regressor ensemble model is summarized as follows:

1. The foreground feature set $\{x_i^{\boldsymbol{o}_\mathrm{F}}\}_{i=1}^{N\boldsymbol{o}_\mathrm{F}}$ and background feature set $\{x_i^{\boldsymbol{o}_\mathrm{B}}\}_{i=1}^{N\boldsymbol{o}_\mathrm{B}}$ are constructed based on those pixels specified by manual labeling.

2. The category labels of $\{x_i^{\boldsymbol{o}_\mathrm{F}}\}_{i=1}^{N\boldsymbol{o}_\mathrm{F}}$ and $\{x_i^{\boldsymbol{o}_\mathrm{B}}\}_{i=1}^{N\boldsymbol{o}_\mathrm{B}}$ are estimated by $h_1(x)$ and $h_2(x)$ respectively, and the results are described as $\{y_{\mathrm{f}1i}\}_{i=1}^{N\boldsymbol{o}_\mathrm{F}}$, $\{y_{\mathrm{b}1i}\}_{i=1}^{N\boldsymbol{o}_\mathrm{B}}$ and $\{y_{\mathrm{f}2i}\}_{i=1}^{N\boldsymbol{o}_\mathrm{F}}$, $\{y_{\mathrm{b}2i}\}_{i=1}^{N\boldsymbol{o}_\mathrm{B}}$.

3. $\{Y_{\mathrm{f}i},1\}_{i=1}^{N\boldsymbol{o}_\mathrm{F}}$ and $\{Y_{\mathrm{b}i},-1\}_{i=1}^{N\boldsymbol{o}_\mathrm{B}}$ are used as the training sets to train the SVR model, where $Y_{\mathrm{f}i}=\{y_{\mathrm{f}1i}, y_{\mathrm{f}2i}\}$ and $Y_{\mathrm{b}i}=\{y_{\mathrm{b}1i}, y_{\mathrm{b}2i}\}$.

2.3 Ensemble model with regressor boosting

The image segmentation results are strongly influenced by the number and position of manual labels, while detailed manual labeling will bring great inconvenience to users. Furthermore, compared to the feature patterns in images, the manually labeled sample pixels are always scarce. These lead to low performance in generalization and robustness. Inspired by Zhou (2011), a regressor boosting algorithm is proposed to fully obtain and use the region segmentation information contained in unlabeled pixels to assist the training of Reg_EL. This method does not need to extract pixel features from two fully redundant views or to evaluate the confidence level of labeling the unlabeled pixels through 10 times cross-validation.

The two independent views of sufficiency and redundancy in regressor boosting are based on the complementary nature of MARS and TPSR. The regressor boosting method is summarized together with the training of the ensemble model and is described as Algorithm 1. The unlabeled pixels of high labeling confidence are chosen to refine the parameters of $h_1(x)$ and $h_2(x)$. Selecting M' samples from U randomly is primarily to enrich the discriminative feature patterns in the training sets of foreground and background. The proposed regressor boosting algorithm is different from COREG (Zhou and Li, 2007) in mainly two respects. First, the way that COREG validates labeling confidence is to do mean squared error (MSE) validation crossed by one learner for the other. However, our method is based on a pair of independent and complementary regressors. That is to say, the decision is made by two regressors together, which has a stricter de-noising rule. Second, MSE confidence level validation in our regressor boosting is carried out from two independent views of sufficiency and redundancy simultaneously, while cross-validation in COREG is carried out from just one view. Therefore, we conclude that the validation rules of label confidence in our method are more rigorous than those in COREG.

2.4 Image segmentation with GC augmentation

To add spatial affinity to the ensemble regression results, GC is introduced. The cost function of GC is described as follows (Boykov *et al*., 2001; Boykov and Jolly, 2001):

$$E(f)=\lambda\cdot\sum_{p\in I}R_p(f_p)+\sum_{\{p,q\}\in\mathbb{N}}B_{\{p,q\}}\cdot\delta(f_p,f_q). \quad (10)$$

The coefficient specifies a relative importance of the region properties (the first term of the right-hand side of Eq. (10)) versus the boundary properties (the second term of the right-hand side of Eq. (10)). The regional term $R_p(f_p)$ assumes the individual penalties for assigning pixel p to 'foreground' and 'background'. The coefficient $B_{\{p,q\}}$ of the boundary term should be interpreted as a penalty for similarity or discontinuity between p and q. $B_{\{p,q\}}$ is large when pixels p and q are similar and it is close to zero when the two pixels are very different.

Algorithm 1 Regression based ensemble model with regressor boosting (Reg_EL)

1. The foreground feature set $\{x_i^{\boldsymbol{o}_\mathrm{F}}\}_{i=1}^{N_{\boldsymbol{o}_\mathrm{F}}}$ and background feature set $\{x_i^{\boldsymbol{o}_\mathrm{B}}\}_{i=1}^{N_{\boldsymbol{o}_\mathrm{B}}}$ are constructed according to manually labeled sample pixels, and the manually labeled pixel set is denoted as $T=\{\{x_i^{\boldsymbol{o}_\mathrm{F}}\}_{i=1}^{N_{\boldsymbol{o}_\mathrm{F}}},\{x_i^{\boldsymbol{o}_\mathrm{B}}\}_{i=1}^{N_{\boldsymbol{o}_\mathrm{B}}}\}$.

2. The $\{x_i^{\boldsymbol{o}_\mathrm{F}}\}_{i=1}^{N_{\boldsymbol{o}_\mathrm{F}}}$ and $\{x_i^{\boldsymbol{o}_\mathrm{B}}\}_{i=1}^{N_{\boldsymbol{o}_\mathrm{B}}}$ are clustered into k clusters by the k-means method to obtain the foreground pixel feature set $\left\{c_i^{\boldsymbol{o}_F},1\right\}_{i=1}^{k}$ and background pixel feature set $\left\{c_i^{\boldsymbol{o}_B},-1\right\}_{i=1}^{k}$.

3. The MARS regressor h_1 and TPSR regressor h_2 are constructed based on $\left\{c_i^{\boldsymbol{o}_F},1\right\}_1^k$ and $\left\{c_i^{\boldsymbol{o}_B},-1\right\}_{i=1}^{k}$, respectively.

4. Refine the parameters of h_1 and h_2:

(I) Randomly select M' ($M'<M$) unlabeled samples $U'=\{x_1, x_2, \ldots, x_{M'}\}$ from the unlabeled pixel set $U=\{x_1, x_2, \ldots, x_M\}$.

(II) Evaluate the labeling confidence of $x_i\in U'$, if it can be used as manually labeled pixels to refine h_1 and h_2:

(i) For $x_i\in U'$, calculate $h_1(x_i)$ and $h_2(x_i)$ separately and evaluate whether $h_1(x_i)$ and $h_2(x_i)$ belong to the same label. If they do, turn to (ii); if not, carry out the process on x_{i+1} until all the elements in U' are taken out.

(ii) If $L_{x_i}=1$, cluster x_i into $\left\{c_i^{\boldsymbol{o}_F},1\right\}_1^k$; if $L_{x_i}=-1$, cluster x_i into $\left\{c_i^{\boldsymbol{o}_B},-1\right\}_{i=1}^{k}$ and record all the elements $\{x_i\}$ in which cluster set x_i is clustered, denoted as $V=\{\{x_i\}, L_{x_i}\}\cup\{c_i^{\boldsymbol{o}_\mathrm{F}}, 1\}_{i=1}^{k}\cup\{c_i^{\boldsymbol{o}_\mathrm{B}}, -1\}_{i=1}^{k}$. During the clustering process, if x_i cannot be clustered to the existing clusters according to the KNN rule, then all the elements in T_{aug} and x_i will be re-clustered into k categories again.

(iii) Retrain the MARS regressor h_1 and TPSR regressor h_2 with $\left\{c_i^{\boldsymbol{o}_F},-1\right\}_{i=1}^{k}$ and $\left\{c_i^{\boldsymbol{o}_B},-1\right\}_{i=1}^{k}$ obtained in the above step.

(iv) Evaluate the labeling confidence of $x_i\in U'$ by the function of $\mathrm{MSE}(h_l)=\sum_{i=1}^{2k+N(c_i^l)}\left\{[y_i-h_l(x_i)]^2-[y_i-h_l^*(x_i)]^2\right\}\geq 0$, where the inputs of $h_1(x)$, $h_2(x)$ and $h_1^*(x)$, $h_2^*(x)$ are the centers of all $2k$ clusters and all elements of the cluster in which x_i is clustered. If the condition is satisfied, $h_1(x)$ and $h_2(x)$ will be replaced with $h_1^*(x)$ and $h_2^*(x)$ and the pixel x_i will be added to the augmented labeled pixel set $T_{\mathrm{aug}}=T\cup\{x_i\}$. Otherwise, refuse the replacement and choose x_{i+1} from U' to keep carrying out step (I), until all the elements of U' are taken out.

5. The regressor boosting process ends. According to the SVR based ensemble method, build the final image segmentation model Reg_EL as $H(x)=S(h_1(x), h_2(x))$.

In former algorithms, the regional penalties $R_p(f_p)$ are set as the negative log-likelihoods of the intensity histograms of manually marked 'foreground' and 'background' pixels. The regressor ensemble results can characterize the probability of assigning 'foreground' and 'background' labels to pixel p. Therefore, the individual penalties for assigning pixel p to 'foreground' and 'background' can be directly set as the normalized results of Reg_EL.

To set the boundary penalties, the ad-hoc function is introduced (Boykov *et al*., 2001):

$$B_{\{p,q\}}=\exp\left\{-\frac{[H(p)-H(q)]^2}{2\sigma^2}\right\}\cdot\frac{1}{\mathrm{dist}(p,q)},\quad(11)$$

where σ can be estimated as the 'camera noise', and $H(\cdot)$ is the regression result of SVR.

The energy minimization problem (10) can be related to the max-flow min-cut theorem, and the minimization result is actually the minimal cut of the graph, which is also the corresponding result of image segmentation.

3 Experiments

Two widely used natural image datasets, the Berkeley segmentation data set BSDS500 (Martin *et al*., 2001) and the Pascal Visual Object Classes (Pascal VOC) (Everingham *et al*., 2010), have been chosen for the experiments. The selection of pixel features is based on experiment, and we have found that the segmentation results in the CIELab color space are slightly better than those in the RGB color space because of the influence of uneven illumination. It is difficult to accurately segment the images with a lot of similar areas existing in both the foreground and background by just taking the three channels of the CIELab color space as the pixel features. After adding the space coordinates of pixels, we have found that when the foreground target is wrapped around by manual labels, it is easy to eliminate the interference of tiny image areas in the background which are similar to those in the foreground. As for the Gabor texture features (Zhang *et al*., 2016), when one scale or orientation Gabor feature is added into the pixels' feature vector, the image segmentation results almost change towards the deterioration direction. Therefore,

the Garbor texture feature is not taken into consideration in this study. In conclusion, the features of any pixel in the image to be segmented are described by a five-dimensional feature vector (three channels of CIELab and coordinates).

To objectively evaluate the performance of the proposed method in natural scene image segmentation, there is neither filtering nor post-processing method adopted to filter out those wrongly segmented tiny areas in the segmentation results.

3.1 Complementarity verification between the MARS regressor and TPSR regressor

To validate the complementary nature existing between MARS and TPSR, a fair comparison experiment has been conducted under the same input manual labels. From Fig. 3, it is clear that those images which are difficult to segment for MARS can be well segmented by TPSR, such as the segmentation of the bird, rock, and butterfly. At the same time, there are some other images which are difficult to segment for TPSR but can be well segmented by MARS, such as the segmentation of cars, houses, and historic buildings. Even for an image that cannot be well segmented by both MARS and TPSR, there are few overlapping regions which are simultaneously wrongly segmented by both MARS and TPSR. Since the training samples are specified by manual labels, the performances of MARS and TPSR will be affected by the number and position of manual labels. We have compared the performance of MARS and TPSR under different label positions and numbers. As can be seen from Fig. 4, there is still obvious complementarity in the segmentation results of MARS and TPSR.

Therefore, we conclude that there are sufficient and redundant natures among MARS and TPSR. The ensemble model by integrating MARS and TPSR is much better than any single model, MARS or TPSR. This is because the probability of two complementary regressors having a large bias at the same time is small. The accuracy of measurement results summarized in Table 1 further demonstrates this.

3.2 Influence of parameter settings on image segmentation

The first parameter in regressor-based ensemble learning is the total number of clusters k in k-means. The purpose of introducing a k-means clustering method to the training sample is mainly to accelerate the training of regressors. Fig. 5 shows the segmentation results of the proposed Reg_EL with different k, and the values of k have been taken as 5, 7, 15, 25, 35, 45, 55, 65, 75, 85, 95, and 105. To more dynamically characterize the influence of k on segmentation accuracy, the accuracy measurement results of TPSR, MARS, and REG_EL with different k's are summarized in Fig. 6. As can be seen from Figs. 5 and 6, when k=75, the optimal segmentation result of the banana image can be obtained, while when k=65, the segmentation result of the llama is almost optimal. It is not hard to see that the parameter k for the optimal segmentation results of different natural images is usually different, which is related mainly to the complexity of the images. However, the result will not always be better with a larger k. As shown in Fig. 6, when k reaches a critical value, the accuracy of image segmentation hardly improves with the increase of k. Furthermore, it will take more time to refine the classifier. In this study, we have found that most natural images can be well segmented when k belongs to the interval from 45 to 85. In the following experiments, k is set as 85.

Another parameter in regressor-based ensemble learning is the M' unlabeled samples which are randomly selected from the unlabeled pixel set $U=\{x_1, x_2, \ldots, x_M\}$ in the procedure of regressor boosting. In a real application, compared to the feature patterns in images, the manually labeled sample pixels are always scarce. It is difficult for a few training samples to cover the key region segmentation information. Therefore, it is meaningful to study the influence of different M' on the segmentation performance. Fig. 7 shows the segmentation results of our method with different M'. In the experiments, the maximum value of M' is limited to 20 000. As shown in Fig. 7, the larger the M', the fewer the tiny areas misclassified. Consequently, to obtain better segmentation performance, more unlabeled pixels should be added to the labeled pixel sample set to refine the classification model. However, there is a balance between accuracy and time taken. With the increase of the number of unlabeled pixels, the training time will get longer. Table 2 further shows the accuracy measurement results of the three images in Fig. 7. In the following experiments, M' is set as 8000.

3.3 Processing of similar areas in foreground and background with different algorithms

In this study, the tiny areas existing in background which share similar colors in the foreground are considered noises. To analyze the noise-resistant ability of the proposed regressor-based ensemble learning method and its competitors, some typical images that share similar colors in foreground and background have been chosen to conduct comparisons. GC (Boykov and Jolly, 2001; Zhou *et al*., 2014), SVM (Wang *et al*., 2011b), and GSC (Gulshan *et al*., 2010) methods have been chosen as representative classical methods for comparison.



Fig. 3 The complementarity performance of MARS and TPSR in interactive natural image segmentation: (a) and (f) are input images with manual labeling; (b)–(e) and (g)–(j) are the results of Reg_EL, MARS, TPSR, and the truth image, respectively

The green and red lines represent the manually labeled foreground and background pixels, respectively. References to color refer to the online version of this figure

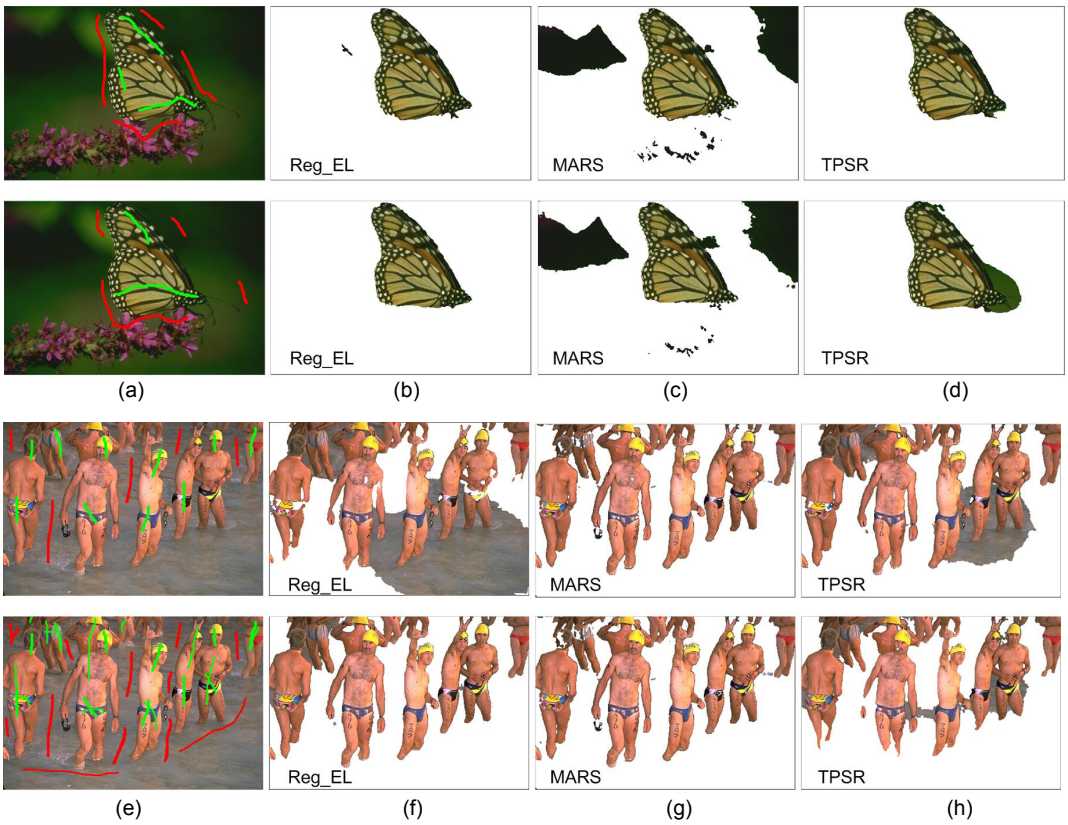


Fig. 4 The segmentation results with different manual labeling: (a) and (e) are input images with manual labeling; (b)–(d) and (f)–(h) are the results of Reg_EL, MARS, and TPSR, respectively

The first row is obtained with different label positions, and the second row is under a different number of labels

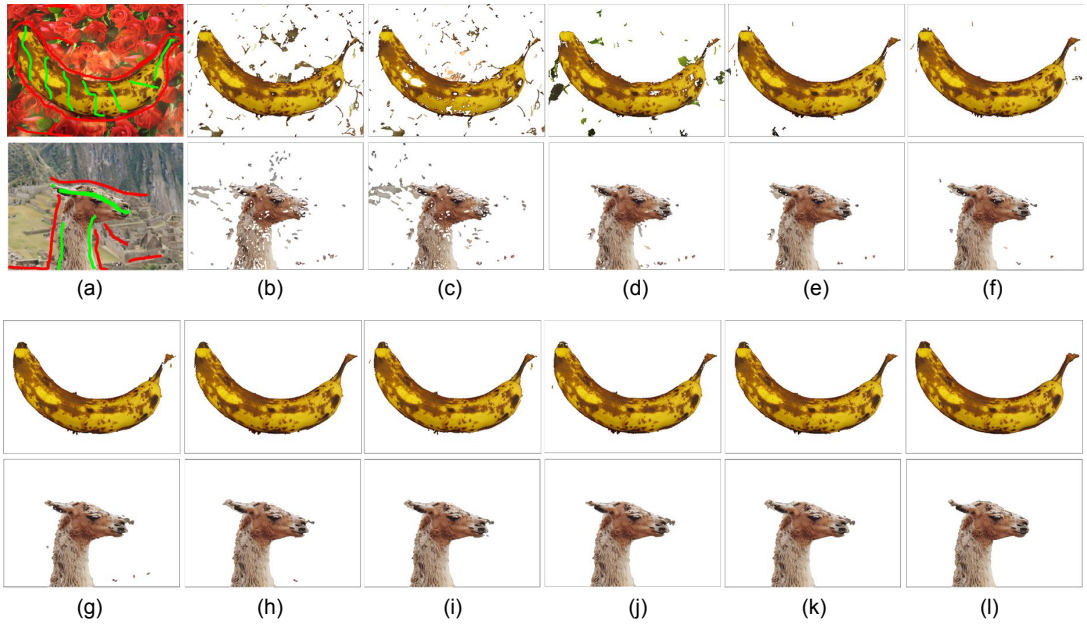


Fig. 5 The segmentation results of Reg_EL with different cluster number *k*: (a) input image with manual labeling; (b)–(k) segmentation results with *k*=5, 7, 15, 25, 35, 45, 65, 75, 85, 105, respectively; (l) truth image

The green and red lines represent the manually labeled foreground and background pixels, respectively. References to color refer to the online version of this figure

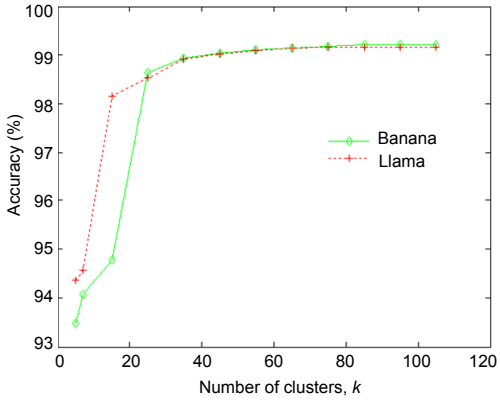


Fig. 6 Accuracy measurement results of Fig. 5

What can be easily seen from Fig. 8 is that the noise-resistant ability of MARS is stronger than that of SVM (Wang *et al*., 2011b), but weaker than that of TPSR. Because of the lack of emphasis on the structure information of pixel features, SVM almost cannot differentiate the foreground and background. There are many details missed in the segmentation result of GC. This is consistent with previous studies (Yang *et al*., 2010; Wu *et al*., 2014). Because of the non-linear ensemble of MARS and TPSR, the performance of the proposed Reg_EL is better than that of former classical algorithms. By adding a boundary constraint to the regression results, the Reg_EL_GC method obtains some improvement compared to Reg_EL and ranks the best of all competitors as summarized in Table 3.

Table 1 Segmentation accuracy measurement results of the six images in Fig. 3

Tested image	Segmentation accuracy (%)		
	TPSR	MARS	Reg_EL
1	96.32	97.80	98.56
2	98.75	98.36	98.86
3	96.53	95.86	97.04
4	93.86	96.87	96.96
5	95.30	98.07	98.52
6	98.67	79.45	99.16

Table 2 Segmentation accuracy measurement results of the three images in Fig. 7

M'	Segmentation accuracy (%)		
	1	2	3
0	78.29	97.43	98.86
4000	98.60	97.92	99.03
8000	98.65	98.04	99.15
12000	98.68	98.06	99.23
16000	98.68	98.11	99.25
20000	98.68	98.12	99.25

3.4 Performance of the proposed algorithm with few or uneven area labels

The result of interactive image segmentation depends on the labeling of manual labels from users, and different numbers or positions of labels may generate a segmentation result seriously inconsistent with the ground truth. Fig. 9 shows the segmentation

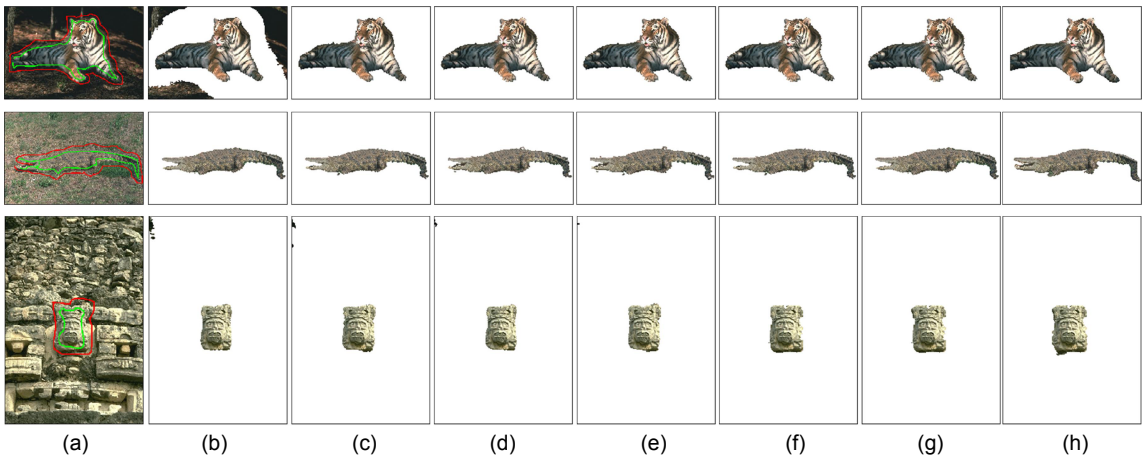


Fig. 7 The segmentation results of Reg_EL with different M': (a) input image with manual labeling; (b)–(g) segmentation results with M'=0, 4000, 8000, 12000, 16000, 20000, respectively; (h) truth image

The green and red lines represent the manually labeled foreground and background pixels, respectively. References to color refer to the online version of this figure



Fig. 8 The processing of a mass of similar areas in foreground and background with different algorithms: (a)–(i) represent the input image with manual labeling, truth image, MARS, TPSR, SVM, GSC, GraphCut, Reg_EL, and Reg_EL_GC, respectively

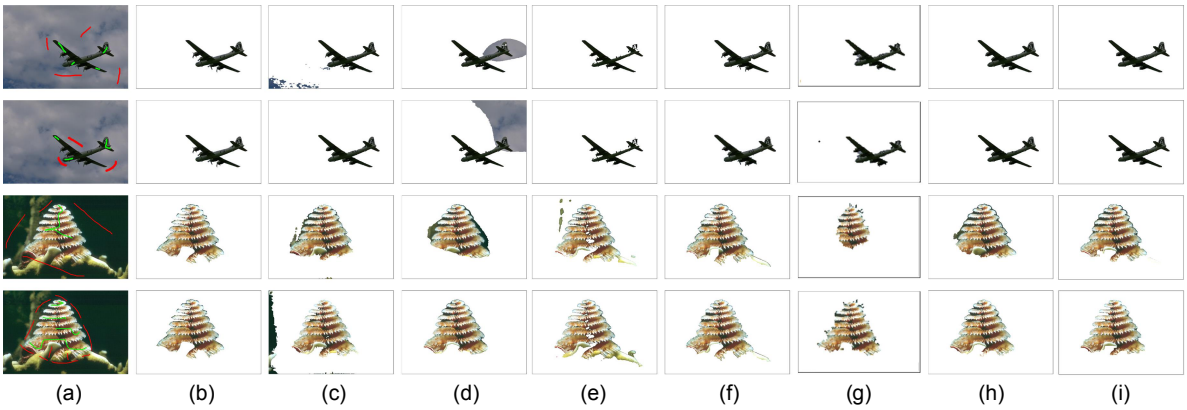


Fig. 9 The segmentation results with few or uneven labels: (a)–(i) represent the input image with manual labeling, truth image, MARS, TPSR, SVM, GSC, GraphCut, Reg_EL, and Reg_EL_GC, respectively

Table 3 Segmentation accuracy measurement results of the four images in Fig. 8

Method	Segmentation accuracy (%)			
	1	2	3	4
SVM	86.38	68.97	87.52	91.90
TPSR	98.93	97.92	98.75	99.35
MARS	92.37	96.47	98.37	99.28
GSC	97.38	97.33	97.97	99.70
GraphCut	97.53	97.17	97.53	99.38
Reg_EL	98.79	98.04	98.75	99.65
Reg_EL_GC	98.81	98.14	98.76	99.67

results of different interactive image segmentation algorithms with the lack of labels or imbalance in area labels. The segmentation accuracy of different segmentation methods has also been summarized to quantitatively compare the segmentation performance on each image. Among all these methods, the performance of Reg_EL_GC is the best (Table 4). The selected GC (Zhou *et al*., 2014), SVM (Wang *et al*., 2011b), and GSC (Gulshan *et al*., 2010) methods are used for comparison.

3.5 Comparison with GC constructed on linearly integrated regressors

To demonstrate the importance of a non-linear SVR fusion method in the proposed Reg_EL_GC method, we conduct some comparisons with a comparative GC model constructed based on the linear combination of the MARS regressor h_1 and the TPSR regressor h_2. The expectation responses of h_1 and h_2 are

Table 4 Segmentation accuracy measurement results of the four images in Fig. 9

Method	Segmentation accuracy (%)			
	1	2	3	4
SVM	98.45	98.43	95.43	94.50
TPSR	94.31	81.56	93.73	97.41
MARS	97.62	99.33	95.70	92.42
GSC	99.49	99.19	97.58	97.32
GraphCut	98.79	99.03	96.51	89.45
Reg_EL	99.40	99.27	95.72	97.54
Reg_EL_GC	99.46	99.37	97.46	97.77

$$y=\begin{cases}+1, & y\in \boldsymbol{O}_{\mathrm{F}},\\ -1, & y\in \boldsymbol{O}_{\mathrm{B}}.\end{cases} \tag{12}$$

It is obvious that h_1 and h_2 can also characterize the probability of assigning 'foreground' and 'background' labels to pixel p. Therefore, in the comparative GC model, the individual penalties for assigning pixel p to 'foreground' and 'background' can be set as the maximum positive responses and minimum negative responses of h_1 and h_2, respectively.

Similar to the proposed Reg_EL_GC, an ad-hoc function is introduced (Boykov *et al*., 2001) to set the boundary penalties:

$$B_{\{p,q\}}=\exp\left\{-\frac{[H(p)-H(q)]^2}{2\sigma^2}\right\}\cdot\frac{1}{\mathrm{dist}(p,q)}, \tag{13}$$

where $H=[h_1(x)+h_2(x)]/2$, and σ is also estimated as 'camera noise'.

For simplicity, we denote the above constructed comparative model as Reg_GC. Results are shown in Fig. 10, and the accuracy measurements are summarized in Table 5. From Fig. 10d, though the contrast of foreground and background regions in the summarization results of h_1 and h_2 is high, there are some artifact foreground responses. If the fusion results of h_1 and h_2 are taken as thresholds directly, there are inevitably wrongly segmented background regions

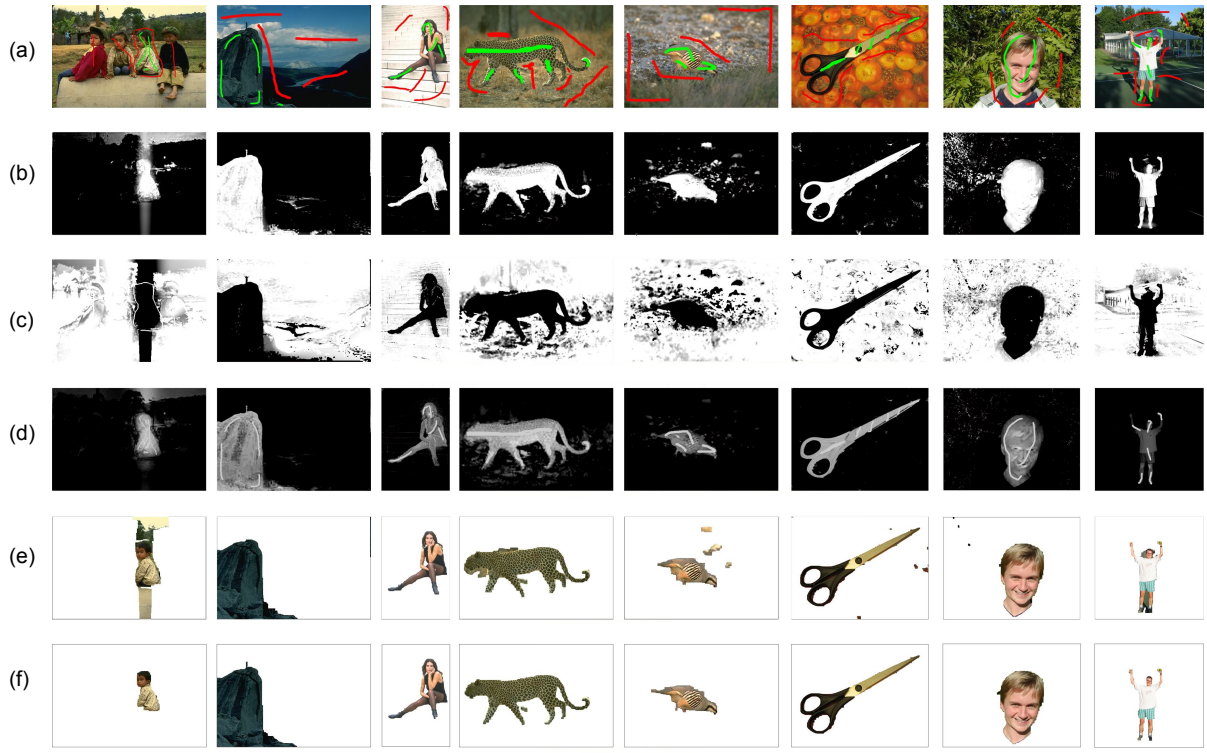


Fig. 10 Segmentation results of the comparative GraphCut based on linear regressor ensemble: (a) input image with manual labeling; (b) regional penalties for assigning the 'foreground' label to the input image; (c) regional penalties for assigning the 'background' label to the input image; (d) images used to compute the coefficient of the boundary term; (e) segmentation results of this comparative model Reg_GC; (f) segmentation results of our Reg_EL_GC

due to no spatial affinity constraint. GC combines the region probability and boundary affinity. Therefore, a lot of artifacts have been lessened in the segmentation results as shown in Fig. 10e. Because the relationship between the regression results of h_1 and h_2 is non-linear, learning based SVR outperforms the linear weighting method in the regressor ensemble. Therefore, the proposed SVR fusion based Reg_EL_GC outperforms linear weighting based Reg_GC.

Table 5 Segmentation accuracy measurement results in Fig. 10

Image	Segmentation accuracy (%)	
	Reg_GC	Reg_EL_GC
1	92.96	99.04
2	99.37	99.71
3	98.89	98.99
4	96.38	97.63
5	97.63	98.98
6	98.67	98.91
7	99.41	99.56
8	98.02	99.01

3.6 Comparison with state-of-the-art methods in different databases

To synthetically evaluate the performance of Reg_EL_GC and compare it with that of the state-of-the-art methods for interactive natural image segmentation, we have made a detailed analysis of the segmentation results relative to the ground truth. These classical algorithms include SVM (Wang *et al*., 2011b), TPSR (Xiang *et al*., 2009), GC (Zhou *et al*., 2014), GrabCut (Rother *et al*., 2004), OneCut (Tang *et al*., 2013), and the geodesic star convexity (GSC) method (Gulshan *et al*., 2010). To observe the improvement by adding a boundary constraint to the regression result, Reg_EL is also included in this comparison. The software packages of these methods have been downloaded from the corresponding author's homepage.

To ensure the consistency of experimental conditions, all algorithms used the same manual labels for the same image. To ensure the segmentation accuracy on all natural scene images, the k value for all natural images used in the experiments has been set to 85. The number M' of chosen unlabeled samples used for REG-boosting training has been set to 8000. To compare the primitive segmentation results of different algorithms, the wrongly segmented tiny areas will not be removed using any post-processing. The ultimate segmentation results with different algorithms are shown in Figs. 11 and 12, which respectively show the segmentation results of natural scene images in the BSDS500 dataset and Pascal VOC dataset. Note that the OneCut method fails in the BSDS500 dataset, and thus the segmentation results of OneCut are not shown in Fig. 11. To quantitatively compare the segmentation accuracy of different algorithms, the performance of each method is summarized in Table 6.

To facilitate the comparison between the image segmentation results and original images, all the foreground areas obtained by segmentation have been replaced by the RGB images of the corresponding position in the original images.

As can be seen from the visual results of image segmentation in Figs. 11 and 12 as well as the objective accuracy measurement results in Table 6, the segmentation results obtained by Reg_EL and Reg_EL_GC are better than those obtained using only the TPSR regressor or MARS regressor. Actually, in the segmentation process, the absolute value of the regression value for most incorrectly segmented areas in the MARS method and TPSR method is near 0, while it is close to or even greater than 1 for the accurately segmented areas. As a consequence, when the results obtained by the MARS and TPSR methods are intelligently combined, the biased small regression value of one regressor will be modified by the larger regression value of the other regressor, so that the final result will be corrected. It is clear that by adding the boundary constraint to the regressor ensemble results, Reg_EL_GC outperforms Reg_EL. The segmentation results of Reg_EL_GC are generally relatively smooth without the isolated wrongly segmented tiny regions.

From Figs. 11 and 12, it can be seen that the interactive image segmentation method based on SVM can obtain a good segmentation result for the images with simple foreground or background, but it is difficult to meet the practical demands for complex images because SVM just emphasizes the maximum interval between sample pixels in the foreground and background. For most natural images, if GC, GrabCut,

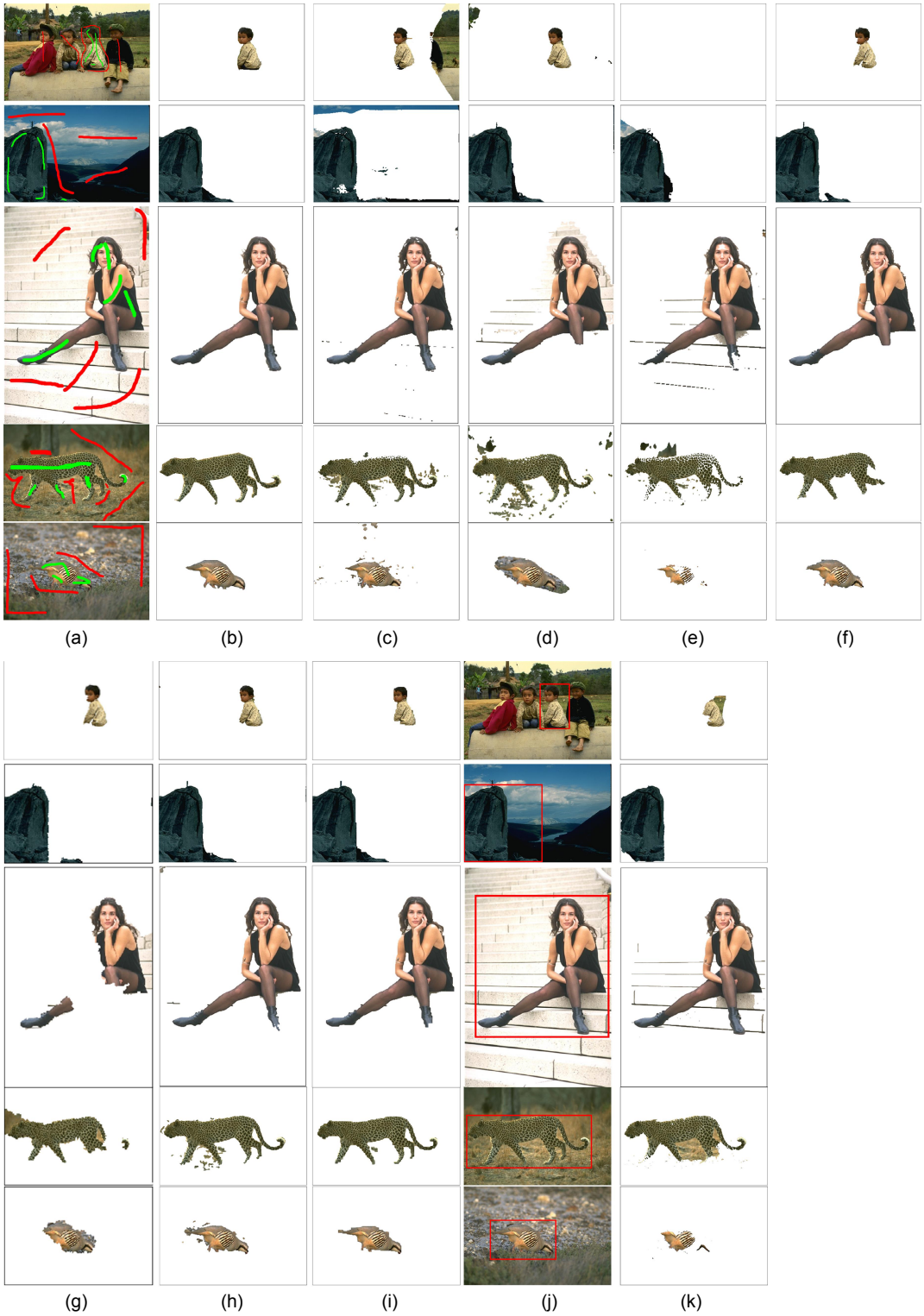


Fig. 11 Comparison of segmentation results of different segmentation algorithms in the BSDS500 database: (a)–(k) represent the input image with manual labeling, truth image, MARS, TPSR, SVM, GSC, GraphCut, Reg_EL, Reg_EL_GC, input image with manual labeling, and GrabCut, respectively

The green and red lines represent the manually labeled foreground and background pixels, respectively. The red rectangle represents the area containing the object obtained using the labeling method of GrabCut. References to color refer to the online version of this figure

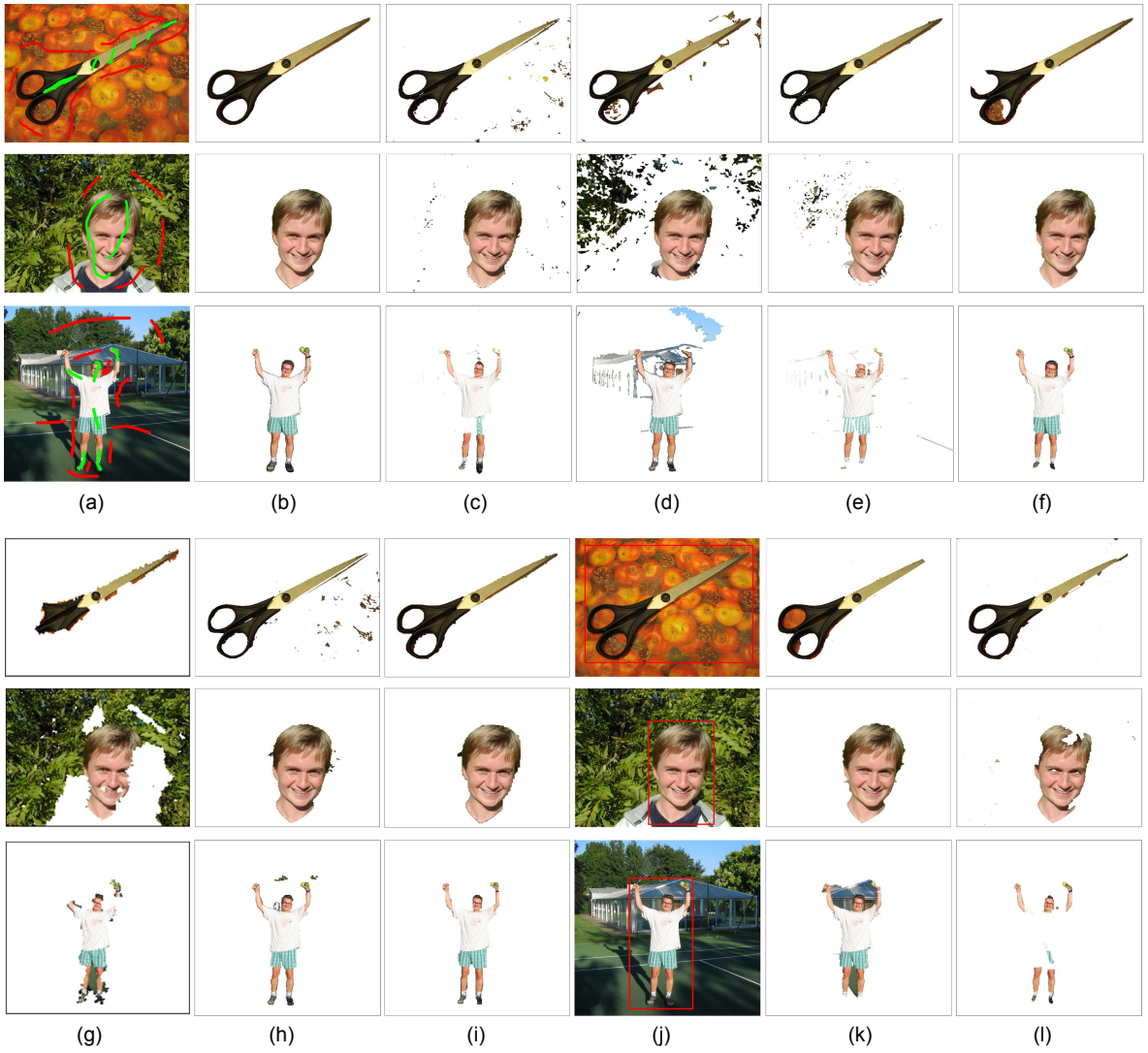


Fig. 12 Comparison of segmentation results of different segmentation algorithms in the Pascal VOC database: (a)–(l) represent the input image with manual labeling, truth image, MARS, TPSR, SVM, GSC, GraphCut, Reg_EL, Reg_EL_GC, the input image with manual labeling, GrabCut, and OneCut, respectively

The green and red lines represent the manually labeled foreground and background pixels, respectively. The red rectangle represents the area containing the object obtained using the labeling method of GrabCut. References to color refer to the online version of this figure

Table 6 Segmentation accuracy measurement results of the eight images in Figs. 11 and 12

Tested image	Segmentation accuracy (%)								
	SVM	TPSR	MARS	GSC	OneCut	GraphCut	GrabCut	Reg_EL	Reg_EL_GC
1	–	97.94	84.32	98.99	–	98.49	96.55	99.02	99.04
2	95.96	99.05	90.47	99.23	–	98.52	97.03	99.53	99.71
3	97.00	89.78	99.00	98.07	–	92.29	96.17	98.97	98.99
4	92.82	92.86	94.82	96.35	–	94.56	93.39	96.86	97.63
5	95.61	96.32	97.80	99.20	97.34	97.77	96.12	98.56	98.98
6	97.11	97.32	97.35	97.63	98.38	95.81	96.46	98.88	98.91
7	97.85	81.74	99.09	99.64	93.73	35.25	99.51	99.51	99.56
8	97.28	93.20	97.45	98.97	96.48	97.31	95.68	98.93	99.01

or OneCut can obtain a good segmentation result, can the proposed Reg_EL and Reg_EL_GC. The segmentation results obtained by GC, GrabCut, and OneCut methods are generally relatively smooth without the isolated wrongly segmented tiny regions. This is mainly due to the fact that these methods use structure information of pixel features combined with the similarity between the pixels to be segmented and the sample pixels. However, for the images with similar foreground and background, it is easy to produce a large segmentation error. Among all these state-of-the-art methods, the GSC method slightly outperforms Reg_EL_GC in several images, but the overall performance is consistent in the two methods. The proposed regressor ensemble model, based on two complementary regressors, not only makes the most of classification information of the manually labeled pixels, but also fully uses the classification information of the unlabeled pixels by the regressor boosting algorithm. As a result, under the same manual labels, the proposed methods can achieve a more accurate segmentation result for those images with the complicated shapes and rich surface details compared to the competitive methods. Therefore, we conclude that the Reg_EL_GC method proposed in this study is comparable with the state-of-the-art methods.

4 Conclusions

Since it is difficult for conventional interactive image segmentation methods to achieve the fine segmentation of those images with complicated and changeable objective shapes, rich surface details, and complex background, a new interactive image segmentation method with regression-based ensemble learning paradigm was proposed in this study. To make full use of the limited manually labeled pixel samples and fully consider the region segmentation information contained in a large number of unlabeled pixels, a new semi-supervised learning method named 'regressor boosting' was proposed. The MARS regressor which can fully use the limited labeled sample data information and the TPSR regressor, relatively insensitive to noise, were integrated using the SVR fusion method to integrate their respective advantages. Because per-pixel regression based image segmentation lacks a spatial affinity constraint, GC was introduced and combined with the regression results of SVR ensemble outputs. The results of extensive experiments on benchmark datasets of BSDS500 and Pascal VOC demonstrated the effectiveness of our method and validated that Reg_EL_GC outperforms state-of-the-art methods for interactive natural image segmentation.

References

Adamowski, J., Chan, H.F., Prasher, S.O., *et al*., 2012. Comparison of multivariate adaptive regression splines with coupled wavelet transform artificial neural networks for runoff forecasting in Himalayan micro-watersheds with limited data. *J. Hydroinform.*, **14**(3):731-744. https://doi.org/10.2166/hydro.2011.044

Balcan, M.F., Blum, A., Yang, K., 2004. Co-training and expansion: towards bridging theory and practice. 17th Int. Conf. on Neural Information Processing Systems, p.89-96.

Blum, A., Mitchell, T., 1998. Combining labeled and unlabeled data with co-training. 11th Annual Conf. on Computational Learning Theory, p.92-100. https://doi.org/10.1145/279943.279962

Boykov, Y.Y., Jolly, M.P., 2001. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. 8th IEEE Int. Conf. on Computer Vision, p.105-112. https://doi.org/10.1109/ICCV.2001.937505

Boykov, Y.Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. *IEEE Trans. Patt. Anal. Mach. Intell.*, **23**(11):1222-1239. https://doi.org/10.1109/ICCV.1999.791245

Ding, J.J., Lin, C.J., Lu, I.F., *et al*., 2015. Real-time interactive image segmentation using improved superpixels. IEEE Int. Conf. on Digital Signal Processing, p.740-744. https://doi.org/10.1109/ICDSP.2015.7251974

Everingham, M., van Gool, L., Williams, C.K., *et al*., 2010. The Pascal Visual Object Classes (VOC) challenge. *Int. J. Comput. Vis.*, **88**(2):303-338. https://doi.org/10.1007/s11263-009-0275-4

Friedman, J.H., 1991. Multivariate adaptive regression splines. *Ann. Statist.*, **19**(1):1-67.

Fu, Z., Wang, L., Zhang, D., 2014. An improved multi-label classification ensemble learning algorithm. *In*: Li, S., Liu, C., Wang, Y. (Eds.), Pattern Recognition. Springer Berlin Heidelberg, p.243-252. https://doi.org/10.1007/978-3-662-45646-0_25

Galar, M., Fernandez, A., Barrenechea, E., *et al*., 2012. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. *IEEE Trans. Syst. Man Cybern. C*, **42**(4):463-484. https://doi.org/10.1109/TSMCC.2011.2161285

Ge, L., Ju, R., Ren, T., *et al*., 2015. Interactive RGB-D image segmentation using hierarchical graph cut and geodesic distance. *In*: Ho, Y.S., Sang, J., Ro, Y.M., *et al*. (Eds.),

Advances in Multimedia Information Processing. Springer International Publishing, p.114-124. https://doi.org/10.1007/978-3-319-24075-6_12

Gulshan, V., Rother, C., Criminisi, A., *et al*., 2010. Geodesic star convexity for interactive image segmentation. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, p.3129-3136. https://doi.org/10.1109/CVPR.2010.5540073

Jian, M., Jung, C., 2016. Interactive image segmentation using adaptive constraint propagation. *IEEE Trans. Image Process.*, **25**(3):1301-1311. https://doi.org/10.1109/TIP.2016.2518480

Jobst, A.M., Kingston, D.G., Cullen, N.J., *et al*., 2016. Combining thin-plate spline interpolation with a lapse rate model to produce daily air temperature estimates in a data-sparse alpine catchment. *Int. J. Climatol.*, **37**(1):214-229. https://doi.org/10.1002/joc.4699

Jung, C., Jian, M., Liu, J., *et al*., 2014. Interactive image segmentation via kernel propagation. *Patt. Recogn.*, **47**(8): 2745-2755. https://doi.org/10.1016/j.patcog.2014.02.010

Kolmogorov, V., Zabih, R., 2004. What energy functions can be minimized via graph cuts? *IEEE Trans. Patt. Anal. Mach. Intell.*, **26**(2):147-159. https://doi.org/10.1109/TPAMI.2004.1262177

Lazaridis, A., Mporas, I., Ganchev, T., *et al*., 2011. Support vector regression fusion scheme in phone duration modeling. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, p.4732-4735. https://doi.org/10.1109/ICASSP.2011.5947412

Lee, Y.S., Cho, S.B., 2014. Activity recognition with Android phone using mixture-of-experts co-trained with labeled and unlabeled data. *Neurocomputing*, **126**:106-115. https://doi.org/10.1016/j.neucom.2013.05.044

Li, Y., Sun, J., Tang, C.K., *et al*., 2004. Lazy snapping. *ACM Trans. Graph.*, **23**(3):303-308. https://doi.org/10.1145/1015706.1015719

Liu, Y., Yu, Y., 2012. Interactive image segmentation based on level sets of probabilities. *IEEE Trans. Visual. Comput. Graph.*, **18**(2):202-213. https://doi.org/10.1109/TVCG.2011.77

Martin, D., Fowlkes, C., Tal, D., *et al*., 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. 8th IEEE Int. Conf. on Computer Vision, p.416-423. https://doi.org/10.1109/ICCV.2001.937655

Menon, R., Bhat, G., Saade, G.R., *et al*., 2014. Multivariate adaptive regression splines analysis to predict biomarkers of spontaneous preterm birth. *Acta Obstetr. Gynecol. Scandinav.*, **93**(4):382-391. https://doi.org/10.1111/aogs.12344

Nguyen, T.N.A., Cai, J., Zhang, J., *et al*., 2012. Robust interactive image segmentation using convex active contours. *IEEE Trans. Image Process.*, **21**(8):3734-3743. https://doi.org/10.1109/TIP.2012.2191566

Ning, J., Zhang, L., Zhang, D., *et al*., 2010. Interactive image segmentation by maximal similarity based region merging. *Patt. Recogn.*, **43**(2):445-456. https://doi.org/10.1016/j.patcog.2009.03.004

Opitz, D., Maclin, R., 1999. Popular ensemble methods: an empirical study. *J. Artif. Intell. Res.*, **11**:169-198. https://doi.org/10.1613/jair.614

Pauchard, Y., Fitze, T., Browarnik, D., *et al*., 2016. Interactive graph-cut segmentation for fast creation of finite element models from clinical CT data for hip fracture prediction. *Comput. Methods Biomech. Biomed. Eng.*, **19**(16):1693-1703. https://doi.org/10.1080/10255842.2016.1181173

Peng, B., Zhang, L., Zhang, D., 2013. A survey of graph theoretical approaches to image segmentation. *Patt. Recogn.*, **46**(3):1020-1038. https://doi.org/10.1016/j.patcog.2012.09.015

Qin, C., Zhang, G., Zhou, Y., *et al*., 2014. Integration of the saliency-based seed extraction and random walks for image segmentation. *Neurocomputing*, **129**:378-391. https://doi.org/10.1016/j.neucom.2013.09.021

Rother, C., Kolmogorov, V., Blake, A., 2004. GrabCut: interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, **23**(3):309-314. http://dx.doi.org/10.1145/1015706.1015720

Shahshahani, B.M., Landgrebe, D.A., 1994. The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon. *IEEE Trans. Geosci. Remote Sens.*, **32**(5):1087-1095. https://doi.org/10.1109/36.312897

Tang, M., Gorelick, L., Veksler, O., *et al*., 2013. GrabCut in one cut. IEEE Int. Conf. on Computer Vision, p.1769-1776. https://doi.org/10.1109/ICCV.2013.222

Wang, T., Sun, Q., Ji, Z., *et al*., 2016. Multi-layer graph constraints for interactive image segmentation via game theory. *Patt. Recogn.*, **55**:28-44. https://doi.org/10.1016/j.patcog.2016.01.018

Wang, X.Y., Wang, Q.Y., Yang, H.Y., *et al*., 2011a. Color image segmentation using automatic pixel classification with support vector machine. *Neurocomputing*, **74**(18): 3898-3911. https://doi.org/10.1016/j.neucom.2011.08.004

Wang, X.Y., Wang, T., Bu, J., 2011b. Color image segmentation using pixel wise support vector machine classification. *Patt. Recogn.*, **44**(4):777-787. https://doi.org/10.1016/j.patcog.2010.08.008

Wu, J., Zhao, Y., Zhu, J.Y., *et al*., 2014. MILCut: a sweeping line multiple instance learning paradigm for interactive image segmentation. IEEE Conf. on Computer Vision and Pattern Recognition, p.256-263. https://doi.org/10.1109/CVPR.2014.40

Xiang, S., Nie, F., Zhang, C., *et al*., 2009. Interactive natural image segmentation via spline regression. *IEEE Trans. Image Process.*, **18**(7):1623-1632. https://doi.org/10.1109/TIP.2009.2018570

Xiang, S., Nie, F., Zhang, C., 2010. Semi-supervised classification via local spline regression. *IEEE Trans. Patt. Anal. Mach. Intell.*, **32**(11):2039-2053. https://doi.org/10.1109/TPAMI.2010.35

Yang, W., Cai, J., Zheng, J., *et al*., 2010. User-friendly interactive image segmentation through unified combinatorial user inputs. *IEEE Trans. Image Process.*, **19**(9):2470-2479. https://doi.org/10.1109/TIP.2010.2048611

Zhang, J., Tang, Z., Liu, J., *et al*., 2016. Recognition of flotation working conditions through froth image statistical modeling for performance monitoring. *Miner. Eng.*, **86**: 116-129. https://doi.org/10.1016/j.mineng.2015.12.008

Zhang, W., Goh, A.T., 2016. Evaluating seismic liquefaction potential using multivariate adaptive regression splines and logistic regression. *Geomech. Eng.*, **10**(3):269-284. https://doi.org/10.12989/gae.2016.10.3.269

Zhang, Y., Song, H., Gu, J., *et al*., 2010. Interactive object extraction using hierarchical graph cuts. Int. Conf. on Audio Language and Image Processing, p.851-858. https://doi.org/10.1109/ICALIP.2010.5685212

Zhang, Y., Wen, J., Wang, X., *et al*., 2014. Semi-supervised learning combining co-training with active learning. *Expert Syst. Appl.*, **41**(5):2372-2378. https://doi.org/10.1016/j.eswa.2013.09.035

Zhou, W., Garcia, E.V., 2016. Nuclear image-guided approaches for cardiac resynchronization therapy (CRT). *Curr. Cardiol. Rep.*, **18**(1):1-11. https://doi.org/10.1007/s11886-015-0687-4

Zhou, W., Hou, X., Piccinelli, M., *et al*., 2014. 3D fusion of LV venous anatomy on fluoroscopy venograms with epicardial surface on SPECT myocardial perfusion images for guiding CRT LV lead placement. *JACC Cardiov. Imag.*, **7**(12):1239-1248. https://doi.org/10.1016/j.jcmg.2014.09.002

Zhou, Z.H., 2011. When semi-supervised learning meets ensemble learning. *Front. Electr. Electron. Eng. China*, **6**(1): 6-16. https://doi.org/10.1007/978-3-642-02326-2_53

Zhou, Z.H., Li, M., 2005. Semi-supervised regression with co-training. 19th Int. Joint Conf. on Artificial Intelligence, p.908-913.

Zhou, Z.H., Li, M., 2007. Semisupervised regression with cotraining-style algorithms. *IEEE Trans. Knowl. Data Eng.*, **19**(11):1479-1493. https://doi.org/10.1109/TKDE.2007.190644