An Overview of Contour Detection Approaches

6 authors, including:

An Overview of Contour Detection Approaches

Xin-Yi Gong[1,2]   Hu Su[1,2]   De Xu[1,2,3]   Zheng-Tao Zhang[1,2]

Fei Shen[1,2]   Hua-Bin Yang[1,2]

[1] Research Center of Precision Sensing and Control, Institute of Automation, Chinese Academy of Science, Beijing 100190, China

[2] University of Chinese Academy of Science, Beijing 100049, China

[3] Tianjin Intelligent Technology Institute of CASIA Co., Ltd, Tianjin 300300, China

Abstract: Object contour plays an important role in fields such as semantic segmentation and image classification. However, the extraction of contour is a difficult task, especially when the contour is incomplete or unclosed. In this paper, existing contour detection approaches are reviewed and roughly divided into three categories: pixel-based, edge-based, and region-based. In addition, since traditional contour detection approaches have achieved a high degree of sophistication, deep convolutional neural networks (DCNNs) have good performance in image recognition, therefore, the DCNNs based contour detection approaches are also covered in this paper. Moreover, the future development of contour detection is analyzed and predicted.

Keywords: Contour detection, contour salience, gestalt principle, contour grouping, active contour.

1 Introduction

Contour is known as a representative feature of imagery object, and thus, its detection is problem in the field of computer vision. Contour detection algorithms are fundamentally required for performing practical tasks, such as object recognition and scene understanding. Until now, numerous researchers have studied the problem constantly, and have gained significant achievements[1].

As described in [2], a contour is defined as an outline representing or bounding the shape or form of an object. Contour detection attempts to extract curves which represent object shapes from images. In fact, the concept of contour is based on human′s common experience, which does not have a formal mathematical definition. Contour is closely related to two additional concepts, i.e., edge[3] and boundary[4], which correspond to the discontinuities in photometrical, geometrical and physical characteristics of objects in images. It is believed that a clear distinction among the three concepts facilitates the feature selection when designing certain special detectors.

In [4], a boundary is described as a contour in the image, representing a change in pixel ownership from one object or surface to another. Some researchers tend to regard contours as the boundaries of interesting regions. However, contour detectors cannot guarantee to produce

closed contours and divide the image into different regions[5]. These cases will frequently occur when contours do not arise from region boundaries. In this sense, the concepts of contour and boundary are closely related, but not identical.

In addition, an edge is represented in the image by changes in the intensity function[3] and can be detected by certain low-level image features such as brightness or color[4]. Therefore, edge detection is a low-level technique which could be applied toward the goal of boundary or contour detection. As illustrated in Fig. 1, the four classical means by which contours are observed are luminance change, texture change, perceptual grouping, and illusory contour, respectively[1]. In the first and second cases, the contours arise from regions boundaries, whereas in the third and fourth cases, global relations give rise to the perception of a contour.

Considering the variety that may emerge from practical applications, contour detection is a difficult task. Generally, the detectors with local features do not achieve satisfactory results, especially for textures, low-contrast objects, or severely noisy images. To pursue a better performance, more sophisticated detectors are developed in which high-level information and sometimes prior knowledge (e.g., the shape of the contour to be detected) are also deployed.

In summary, contour detection is of theoretical and practical significance. Despite all of that, there lack overviews of research studies related to this field. To the best of our knowledge, the most recent overview is published in 2011[1]. It focuses mainly on "edge and line oriented approaches", while other types of approaches are not discussed in depth. Moreover, this review, to some degree,

(a) Luminance changes   (b) Texture changes   (c) Perceptual grouping   (d) Illusory contour
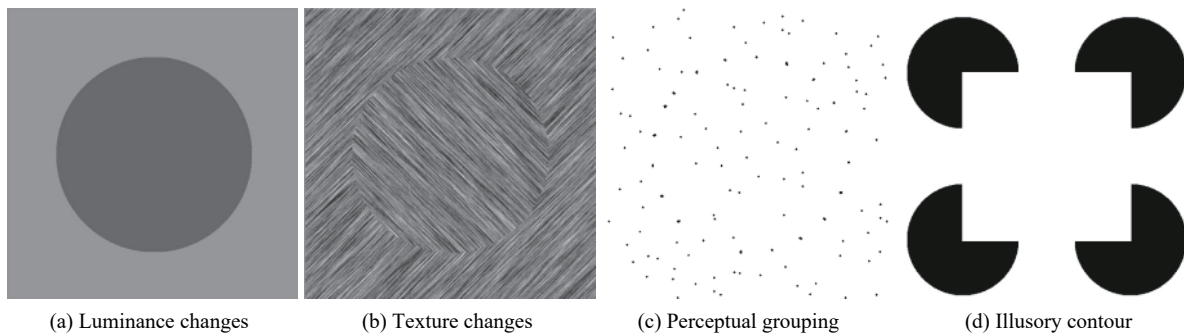
Fig. 1   Four classical means of observing contours

lags behind the development of the era since remarkable progress has been made on this subject. Specially, deep learning techniques[6] are successfully applied to solve the problem of contour detection, achieving comparable performances with the other state-of-the-art methods[7]. Therefore, an overview of the contour detection approaches that includes the recent progresses is necessary.

This paper presents a taxonomy of the existing contour detection algorithms. We attempt to distinguish the following categories. The first one, pixel-based approach, directly determines whether each pixel belongs to a contour according to certain extracted features. The second one, edge-based approach, detects edges and then contours are obtained by grouping or optimizing the edge fragments. The third one, region-based approach, regards contours as boundaries of interesting regions and internal information of the region is taken into account. The final category is deep-network-based approaches, in which deep networks with good generalization abilities are proposed for contour detection.

The rest of this paper is organized as follows. In Section 2, pixel-based approaches are discussed in two steps: feature selection and classification. Edge-based and region-based approaches are discussed in Sections 3 and 4, respectively. In Section 5, recently published papers on deep networks for contour detection are reviewed. Section 6 presents specific applications of contours. Finally, the paper is concluded in Section 7.

2 Pixel-based approaches

In pixel-based approaches, features are constructed and then employed to determine whether each pixel of the image belongs to a contour. Discontinuity in grey-scale intensity is the most salient features in the image and is the first to be discovered. By convolving the image with local filters, this feature can be detected as the pixels with the highest gradient magnitude in their local neighborhood. To extract the discontinuity feature arising from step edge, several linear filters[8–10] are introduced, such as Sobel, Prewitt and Canny operators. However, as Perona and Malik[11] pointed out, real image edges are usually not step functions, but are more typically a combination of steps, peaks, and roof profiles. Step edges are

thus an inadequate model for the discontinuities in the image that result from the composite edges. A class of nonlinear filters, known as quadratic filters, are suggested in [11]. In addition, filters of different scales[12, 13] and orientations[14] are considered as well for providing a richer description.

Nevertheless, the discontinuity feature is not robust, especially in the case of textures and ambiguities. Therefore, it is not appropriate to employ this feature alone to accurately localize an object contour, although it has been widely used in edge detection. In view of this, in recently proposed contour detection approaches, attempts are made to employ higher-level features. Considering the motivations of the features, these approaches can be divided into two categories, namely, approaches based on brain-inspired features and approaches based on natural features, both of which we would discuss below.

2.1 Methods based on brain-inspired feature

Visual neural mechanism forms a theoretical basis for methods based on brain-inspired features that imitate the reaction of human perception in contour detection. Very significant progress has been made in the research of the visual neural mechanism which we now briefly introduce. Firstly, several concepts need to be clarified so as to smooth the following discussion. The primate visual cortex is in general referred to as   In V1, the receptive field (center area) is called as classical receptive field (CRF) and its surrounding area is called non-classical receptive field (NCRF). As identified by the studies[15–19], stimuli falling only in NCRF is not sufficient for driving spiking responses, but can modulate the activity evoked by the stimulus located within the CRF. The modulatory effect is referred to as center-surround interaction. Intense research of cortical neurons in CRF and NCRF has revealed that the visual cortex plays an important role in feature extraction, and that center-surround interaction is desirable for contour perception. Different models are developed to construct the features in accordance with the responses of different areas in V1, which are respectively described in detail as follows.

Daugman et al.[20, 21] describe the response of stimuli in the CRF of simple cells which uses two-dimensional (2D) Gabor filters. The filters fit the 2D spatial response profile of each simple cell in the least-squared error sense (with a simplex algorithm) to estimate whether each pixel is on an edge. In addition, to describe the response of stimuli in the CRF of complex cells, the concept of Gabor energy is proposed in [22]. It is computed by using a bank of orthogonal pairs of Gabor filters for texture analysis and to obtain better results. Zhao et al.[23] calculate the Gabor energy maps and the preferred orientations by convolution with a series of directional Gabor functions. The histograms of the preferred orientation are counted to calculate the modulations and then to ameliorate the Gabor energy. It is demonstrated in these studies that the Gabor filters can approximately simulate the neuronal response characteristics of CRF, such as orientation and frequency. In allusion to the inhibition of the NCRF, various models have been proposed, represented by difference of gaussian (DoG) proposed by Sceniak et al.[24, 25] and ratio of Gaussian (RoG) proposed by Cavanaugh et al.[26]. For the sake of texture suppression in contour detection, Grigorescu et al.[27–29] utilize the NCRF inhibition mechanism, which consists of isotropic and anisotropic inhibition. They consider isolated lines and edges to be non-texture features, which are not affected by the inhibition, while groups of lines and edges viewed as texture features are suppressed. Thereafter, the NCRF inhibition mechanism is further investigated to improve the effect of contour detection. For example, Papari et al.[30, 31] split the inhibition surround into two truncated half-rings oriented along the concerned edge which can reduce the impact of self-inhibition. Ursino and La Cara[32] present a model of contour extraction based on inhibition mechanisms that combine the feed-forward and feedback, the former of which suppresses non-optimal stimuli and ensures contrast invariance and the latter suppresses noise. Long et al.[33–35] propose an adaptive inhibition approach to remove the non-meaningful texture elements for effective contour extraction. Azzopardi et al.[36, 37] propose a computational model of a simple cell that uses as afferent inputs the responses of model lateral geniculate nucleus (LGN) cells with center-surround receptive fields, which is referred to as the combination of receptive fields (CORF) model. In [38], multiple features are suggested for the classical center-surround inhibition which is common to most cortical neurons, such as orientation, luminance, and luminance contrast. For NCRF modulation, in addition to the inhibitory effect, the facilitatory effect, the importance of which is fully validated[39], is also applied to contour detection[40, 41]. Tang et al.[42, 43] establish a unified contour model which fully utilizes the configurational feature of smooth contour to renovate the incomplete contour. The method is appropriate for complex natural scene and more in line with physiological mechanisms. As demonstrated, the NCRF-related methods can dramatically suppress trivial edges arising from texture and thus improve the contour detection performance. For more details on this issue, please refer to [44].

2.2 Methods based on natural features

Approaches based on natural features use the features that are extracted according to natural attributes of the image, such as brightness and color. The approaches include two crucial components in implementation, feature selection and determination.

2.2.1 Feature selection

As mentioned above, texture and noise are important factors that influence the effectiveness of contour detection. To overcome this difficulty, the discontinuity feature is combined with other gradient features, including color and texture gradients, resulting in a probabilistic detector Pb[4]. Motivated by the gradient features in Pb, several related features have been proposed. In [45, 46], color and gradient information are considered via local, textural, oriented, and profile gradient-based features. In [47], additional localization and relative contrast cues are defined and fed to the boundary classifier. The localization cue captures the distance from a pixel to the nearest peak response and the relative contrast cue normalizes the contrasts in two half disks around each pixel. In [48], the hand-designed gradient features in Pb are replaced by the sparse code gradients (SCG) features, bringing in a significant improvement in detection accuracy. SCG features play a similar role with gradient features in measuring local contrast. However, SCG features can automatically learn from image data through sparse coding, thus minimizing human's involvements. Dollar et al.[49] spare more efforts to reduce human burden in designing and selecting appropriate features. In the literature, thousands of simple features computed on images patches are considered to establish a discriminative model. These features include gradients at multiple scales and locations, difference between histograms computed on filter responses at multiple scales and locations, and Haar wavelets as well.

In addition to the above approaches that aim to extract local information, the spectral graph theory, particularly its normalized cuts criterion, provides a means to explore the role of global information in contour detection. The normalized cuts criterion is investigated and applied to image segmentation in [14, 50]. In [5, 51], it extended and improved the performance of detector Pb. As a result, a novel detector gPb is developed which involves both local and global information. In the framework, an affinity matrix W is defined, the elements of which measure the similarity of pixels. Afterwards, diagonal matrix D is calculated with elements $D_{ii} = \Sigma_{ii} W_{ii}$ and the generalized eigenvectors of the system $(D - W)v = \lambda Dv$ are solved. Local contrasts extracted from the eigen-

vectors are combined to provide global information. Despite of its good performance, detector gPb is limited in scalability as it has a high computational cost and memory footprint. In view of this point, an improved normalized cuts algorithm is proposed[52, 53], providing a 20 speed-up to the eigenvector computation.

Therefore, the issue is briefly discussed here. Traditional classification methods[9] are based mainly on dividing the response of filter with a predetermined threshold, or finding local maximum of the responses. In recent research, learning-based methods attract increasingly attention. Compared with traditional methods, the learning-based methods with various image cues can be used to achieve higher robustness. In [4], comparison experiments are conducted with a wide array of classifiers, including hierarchical mixtures of experts (HME), support vector machine (SVM) and so on. Accordingly, logistic model is chosen therein. Consistent with the observations reported in [4], the logistic model is also adopted in [47, 51]. Another classification method, known as probabilistic boosting tree (PBT), is employed in [49] to deal with thousands of features. Until now, we have reviewed some of the existing contour detection methods. We offer a clearer descriptive presentation in Table 1.

Table 1   Summary of the pixel-based contour detection methods

Methods	Image features	Classification	Attribute
Sobel, Prewitt, Gaussian[9], Laplacian[10], Nonlinear filters	Gray gradient	Threshold	Local
Method in [47]	Localization, relative contrast	Logistic	Multiscale, local
Pb[4], SCG[48]	Brightness, color, texture gradients	Logistic	Local
Method in [49]	Histograms of filter responses, gradients, Haar wavelets	Probabilistic boosting tree	Multiscale, local
gPb[5, 51], MCG[52, 53]	Brightness, color, texture gradients, affinity	Logistic	Multiscale, global
Method in [56]	SIFT feature	Threshold	Multiscale, local

2.3 Discussions

In general, it can be concluded that continuous effort has been invested in developing novel features to improve detection accuracy. Different features have been proposed, involving both local and global image information. The features can be combined to improve the robustness, especially when textures and noise occur. The first attempt is reported in [4], which combines brightness, color, and texture gradient features. However, the extent to which the features are sufficient and necessary has not yet completely understood. Dollar et al.[49] attempt to overcome the difficulty, where a large number of simple features are employed to train a discriminative model. The primary advantage of such approach is that the human effort required is minimized, and instead, the work is shifted to classification algorithm. Up to now, few research studies have been conducted on this topic and further discussion is necessary.

For pixel-based approaches, another significant concern is to figure out the range of scales that contours may appear in the images. This issue has been addressed in early research studies on edge detection. In [12, 54], focusing on the Laplacian of Gaussian (LoG) operator, a rigorous analysis of linear edges at different scales is carried out to study the mutual influence of edges. Basic knowledge about the behavior of edges in scale space is provided and can be used for scale space reasoning. In [55], a mechanism is proposed to automatically determine the scale levels that are appropriate for extracting a given edge. To resolve the problem of contour detection, the detection accuracy of several detectors mentioned was further improved by taking the advantage of multi-scale information. Ren[47] demonstrates the benefit of combining information from multiple scales of the Pb detector developed in [4]. Furthermore, a multi-scale version of the Pb detector, named mPb, is proposed in [5, 51]. In [49], gradient features at multiple scales are employed in the discriminative model. However, some aspects of contour detection still have not been investigated sufficiently and more research is needed, such as selecting of the scales of operators for an image and effectively combining of the information recovered at different scales.

3 Edge-based approaches

Edge-based approaches are based on contour related edges or curves provided by edge detectors or human prior experience, aiming to determine whether they are contained in a certain contour. In many of these approaches, global optimization is conducted, whereby information from the entire image can be taken into consideration simultaneously. The difficulty is, in most cases, the optimization problem is NP-hard, which makes it unlikely to find the global optimum in polynomial time. Edge-based approaches can be categorized into two classes: edge grouping and parametric active contour.

3.1 Contour extraction

Although it has not been strictly defined, contour extraction has been mentioned in some of the literatures. In [57], it is regarded as a process of identifying a subset of

fragments produced by preprocessing and connecting the fragments in that subset. In [58], the authors point that contour extraction is a two-step process: edge detection followed by edge grouping. As stated, the grouping process, which is primarily addressed in this section, involves collecting individual edge points together to form continuous contours.

Note that perceptual grouping is a fairly general process that may emerge from different domains. As explained in [59], grouping processes rearrange the given data by eliminating irrelevant data items and sorting the rest into groups, each of which corresponds to a particular object, as illustrated in Fig. 2[60]. In [59], a general grouping approach is proposed, which can be used for grouping various types of measurements, and for incorporating different grouping cues. In our context, measurements are edge elements, and the Gestalt laws which have been used as guidelines for many edge grouping methods, need to be introduced. Gestalt laws are the factors that lead to human perceptual grouping, which can be summarized as similarity, proximity, continuity, symmetry, parallelism, closure and familiarity.
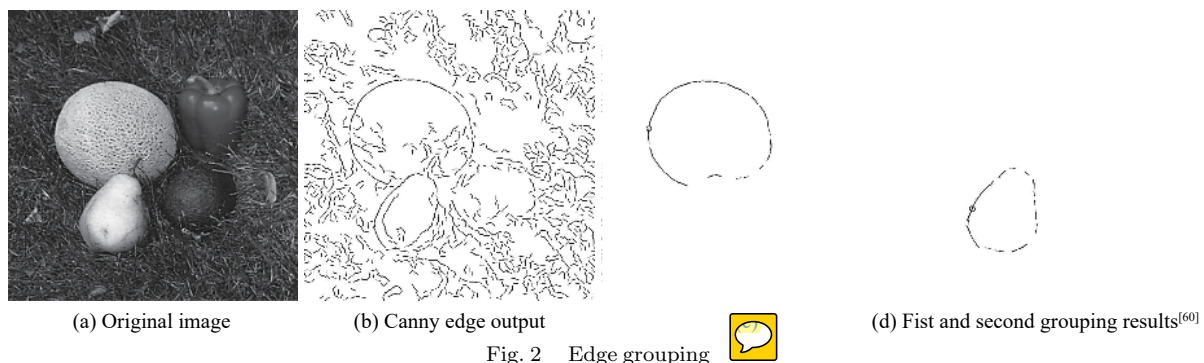

(a) Original image   (b) Canny edge output   (d) Fist and second grouping results[60]

Fig. 2   Edge grouping

Earlier research studies[58, 61, 62] have concentrated on probability modeling which determines the likelihood certain edge element contained in the same contour. These studies differed in the way of probability calculation. More specifically[58], is based on empirical statistics, whereas in [61, 62], the Bayesian inference is adopted. Subsequent work on probability edge grouping focuses on saliency measure, which can be derived by transition probabilities between different edge elements. In [60], the concepts of link saliency and contour saliency are introduced, which are used to identify smooth closed contours, bounding objects of unknown shape in real images. In [60], a method, named stochastic completion fields[63, 64], is adopted for the calculation of transition probability. The method[63, 64] models the Gestalt laws of proximity and continuity by the distribution of smooth curves which are traced by particles that move with constant speed in directions undergoing Brownian motion. The method is not claimed as a saliency measure, but it is easy and natural to use this method to compute saliency. In addition, different saliency measures have been proposed[64–68]. Each of the measures is a function of a set of affinity values assigned to pairs of edges, incorporating the Gestalt principles with good continuation and proximity in some form. However, the affinity between a pair of edges is defined somewhat differently in different measures. Their performances are compared in [69]. As reported in [69], the measure of perceptual saliency significantly affects the performance of the grouping method and thus needs to be emphasized. In addition to grouping edge elements, the saliency measure is proven to be effective in the detection of illusory contours, as shown in Fig. 3.
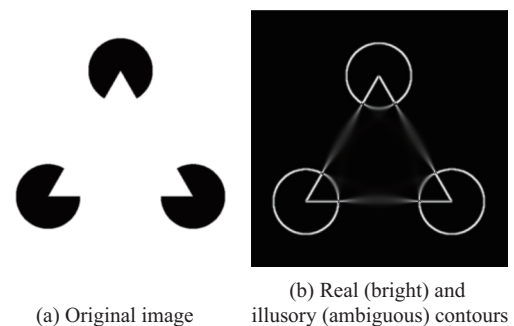

(a) Original image   (b) Real (bright) and illusory (ambiguous) contours

Fig. 3   Detection of illusory contours[63, 64]

The optimization framework is pioneered by Jermyn and Ishikawa[70] and is generalized in [71]. The difference between them is that, the latter imposes curvature regularity of the region boundary whereas the former does not. The authors also provide graph-based methods to solve the optimization problems, which, in most cases, are NP-hard, meaning that it unlikely to find the global optimum in polynomial time. Based on the framework, the approaches called ratio contour are proposed in [57, 72]. It is noteworthy that the above-mentioned methods are designed for extracting salient region boundaries, and in particular, the approach presented in [72] is intended to detect boundaries with symmetry. As the concept of boundary is much narrower than that of contour, their application areas are limited. The restriction can be overcome by using the graph topology proposed in [73]. In this topology, each edgel is represented by two duplicated nodes with opposite directions and the connections[74] between duplicated nodes are added, in which way open

curves could be represented by circles. The same topology is adopted in [75]. Unlike the method in [73], the number of groupings recovered by the algorithm in [75] can be specified by a human user. The application scope is thus expanded. Edgel mentioned here is not a universal concept and has been referred to in some of the above literatures[73, 75, 76] without a strict definition. We take the description in [76] to illustrate the concept:

Edge pixels with very similar gradient direction in a local neighborhood are combined to edgels. The edgel is located at the center position and has the mean direction of the local estimate.

In addition, other grouping algorithms for certain specific purpose are proposed. For example, the path-based grouping approach is proposed in [76], in which connectedness of image elements via mediating elements is sufficiently explored. The grouping approach yields superior results when objects are distributed on low-dimensional extended manifolds in a feature space. In [77], based on the defined cost functions, the proposed grouping approach tends to extract a small set of smooth curves. With the curves, a global interpretation of the image can be derived. And the approach proposed in [78] is designed to find convex sets of line segments. In essence, the optimization approaches mentioned in this paragraph differ mainly in their choice of grouping principles and optimizing approaches. A clearer illustration is summarized in Table 2.

Table 2   Summary of the grouping principles and optimization methods

Grouping objectives	Grouping principles	Optimization methods
General grouping method[59]	Maximum likelihood	Graph-based optimization
Region boundaries[70, 71, 57]	Proximity[70, 71, 57], continuity[71, 57]	Graph-based optimization
Region boundaries[72]	Proximity, symmetry	Graph-based optimization
Closed contours[61]	Maximum likelihood	Graph-based optimization
Salient curves[73]	Cut cost, order, thickness	Graph-based optimization
Salient curves[75]	Response of edge detector, continuity	Graph-based optimization
Salient curves[76]	Connected ness, proximity, smoothness	Iterative optimization
A set of smooth curves[77]	Maximum likelihood	Greedy approximation

3.2 Contour completion

An additional point that plays a critical role in human perception and thus needs to be emphasized is the closure principle. Enforcement of the closure principle can produce more meaningful contours. However, given the noise and ambiguities, contour frequently breaks into fragments, even with state-of-the-art edge detectors. Unlike other relatively local Gestalt principles such as proximity and continuity, the closure principle is more global, and thus the enforcement of such principle is not a simple task. Some attempts to address the issue are reported in [2, 79, 80]. Both of the approaches adopt conditional random field (CRF) model to achieve contour closure. The main differences between them lie in the means used to complete contours across gaps and the potential functions employed in the CRF model. The approaches have the limitation that they can be only applied to contours with small and simple gaps; otherwise, large deviation will occur. A more general gap filling strategy is proposed in [57], which can, to some extent, overcome not only the limitation, but also the negative impact of noise at fragment endpoints. However, there is a possibility that the filled contour would deviate significantly from the original curve. Given this difficulty, the topic still requires more discussions and more efficient approaches.

3.3 Edge-based active contour

Active contour is a variational approach to contour localization, which is also referred to as snake or deformable model in literatures. The main idea is to evolve a hand-drawn curve into the specific line or contour by minimizing predefined energy function, as shown in Fig. 4. The basic concept of active contour is first proposed by Kass et al.[81] in the 1980s. Another pioneer approach that's worth mentioning is to apply level set to solve the model of active contour, which is carried out by Caselles et al.[82–84]. In their approaches, the contour can be expressed implicitly, and thus, its application range is expanded. Built on these works, further research has been conducted and significant progress has been made in recent years. Active contour approaches can be divided into two categories: edge-based approaches and region-based approaches. They are differentiated in whether the luminance inside the curve-bounded region is involved in the definition of energy function. More explicitly, the region-based approaches derive a contour representation from a segmentation of the image into well-defined regions, while in the edge-based approaches, a contour representation fitting to boundary points is characterized by the differential property. Edge-based approaches are reviewed in this section and the region-based approaches

are followed in the next section. With the parametric representation of a snake $v(s) = (x(s),\ y(s))$[81], the energy function is expressed as

$$E_{snake}^* = \int_{snake}^* E_{snake}(v(s))\mathrm{d}(s) = $$
$$\int_0^1 (E_{int}(v(s)) + E_{img}(v(s)) + E_{con}(v(s)))\mathrm{d}s \quad (1)$$

where E_{int}, E_{img} and E_{con} respectively represent different aspects of the forces that drive the curve to the boundary of the object. In particular, E_{int} represents the internal energy of the curve, E_{img} gives rise to the image forces, and E_{con} gives rise to external constraint forces. According to [81], E_{img} can be combined with E_{con} to form the external energy, the value of which represents how closely the curve approximates to the actual contour[83]. By minimizing the energy function, the curve is evolved until it reaches the boundary of the object.

The model proposed in [81] combines the local features with a priori information such as location, shape, and size, making it advantageous in high accuracy and better adaptability. Moreover, it can effectively detect illusory contours. An example is presented in Fig. 4. In the figure, with the initial configuration drawn by the user (Fig. 4 (a)), the model can detect both edges and boundaries (Fig. 4 (b))[85]. However, it also has some drawbacks, such as slow convergence, strong dependence on the initial configuration, and poor performance in detecting low-contrast contours. To overcome the limitations, a large variety of improvements have been developed. For example, balloons snake[86, 87], gradient vector flow (GVF) snake[88], and the model in [89] are developed, all of which aim to reduce the dependence on the initial curve. Different strategies have been adopted in these models. In contrast to the original model proposed in [81], additional forces are introduced to the ballons snake and GVF snake. Futhermore, in these models of [81, 86–88], the curve can move in one direction, whereas in the model in [89], the curve is allowed to evolve with more freedom, in the inward and outward directions simultaneously.
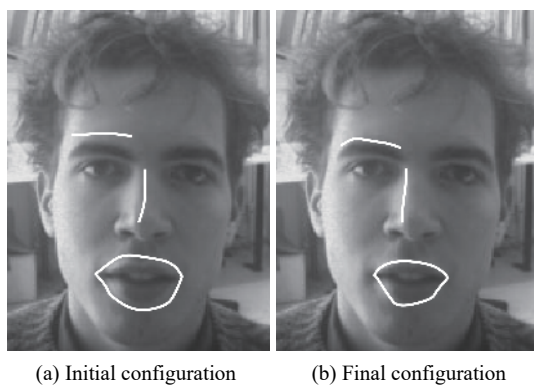

(a) Initial configuration   (b) Final configuration

Fig. 4   Example of active contour[85]

In addition, the geodesic active contour (GAC) model proposed in [83] improves the real-time performance, but having the risk of leaking through the boundary. This risk can be avoided in the model in [90] which takes the shape information into account. In [91], the snakes model is extended to color images. Other examples, including the Fourier snakes[92], the dual snakes[93, 94], the water snake[95], and the directional snakes[96], are also typical modifications of active contour model, which, to some extent, overcome the disadvantages of the original model. For an overview, please refer to [85, 97].

3.4 Discussions

We classify the edge-based approaches into two categories according to whether or not they require an initialization from the user. The first category, discussed in Section 3.1, aims to group edge pixels together into contours according to the Gestalt laws. In the second category, reviewed in Section 3.2, the initial curve is evolved by minimizing an energy function.

In the former category, optimization is usually performed in a graph-based framework. The nodes in the graph have different representations in different optimization methods. The most common representations are image pixels[70], line segments[71], and edgels[73, 75]. In general, the edgel representation may be a more convenient choice when considering the Gestalt principles of closure, proximity, and continuity. In addition, as analyzed in [57], the choice of the most-salient boundary may depend on image scale. In this sense, another consideration should be given to design scale-invariant proximity measures.

In the latter category, active contour is an energy-minimizing approach guided by the external forces and influenced by the image forces that attract the initial drawn curve to the edges and contours. As demonstrated in [98], the approach has a unique advantages of locating the object accurately and efficiently. Moreover, it can effectively detect illusory contours. Nevertheless, original model also has the limitations such as poor real-time performance, strong dependence on the initial configuration and being sensitive to noise. Many modifications have been proposed, to some extent, to overcome those disadvantages. However, we should note that most of them still have their own limitations, and few could satisfy practical requirements. Considering of this, we suggest that, it is necessary to investigate how to remove the requirement of drawing an initial curve and how to improve the robustness.

4 Region-based approaches

Regarding contours as boundaries of interesting regions, region-based approaches take advantage of internal information of the regions to enhance their effectiveness and robustness. The advantages of the approaches are indisputable, but they also have disadvantages, be-

cause they are invalid for contours that do not arise from region boundaries. These approaches are considered in this section from two different aspects: region segmentation approaches and region-based active contours.

4.1 Region segmentation approaches

In fact, region segmentation approaches are closely related to another research area, image segmentation, which plays an important role in object reorganization and scene understanding[99–101]. A detailed discussion on segmentation would not provided here because of space limitation. In this section, only those approaches that have a direct application on contour detection are discussed. A classical region-based approach, named ultrametric contour map (UCM), is proposed in [102, 103], which addresses region-based segmentation and contour extraction as a single task. The key of the approach is the definitions of ultrametric distances, where local contrast and region contribution are involved in measuring the dissimilarities between adjacent regions. Furthermore, the authors proposed an oriented watershed transform (OWT) algorithm in [18] to form initial regions for the construction of UCM. And the study is further extended in [52, 53] by effectively utilizing multiscale information. Ming et al.[104, 105] introduce a contour detection framework that combines region and contour representations based on the winding number concept. In their approach, region similarity and contour cues such as curvature are involved.

4.2 Region-based active contour

As analyzed in Section 3.2, the region-based active contour models compute the sum of integrals over the regions bounded by the curve, rather than over the boundaries. Different from edge-based models, the region-based models can detect boundaries which are not necessarily defined by gradient or not very smooth. Therefore, we know that, region-based models are usually more robust to noise and less sensitive to initialization than edge-based models.

Cohen et al.[106] first suggest that region information should be considered in energy function. According to the energy function in which region function is involved, the evolution speed of the curve could be determined. Subsequently, using Mumfor-Shah functional and level sets, Chan and Vese[107] propose a region-based model which can effectively detect boundaries that are not defined by gradient. The energy function is defined as

$$E = \mu Length(C) + \upsilon Area(inside(C)) +$$
$$\lambda_1 \iint_{inside(C)} |I_{(x,y)} - c_1|^2 \mathrm{d}x\mathrm{d}y +$$
$$\lambda_2 \iint_{outside(C)} |I_{(x,y)} - c_2|^2 \mathrm{d}x\mathrm{d}y \quad (2)$$

where $\mu = 0$, $\upsilon = 0$, and λ_1, $\lambda_2 = 0$, are fixed parameters. For effective computation, level sets methods are utilized in [107] to transform the mixed integrals over regions and over boundaries into the single integrals over boundaries. In [108], the authors then proposed a new multiphase level set framework for image segmentation, which can be seen as the generalization of the model proposed in [107]. However, in the models proposed in [107, 108], only region information is used. Considering that and to further improve the performance, the fusion methods which are based on both region information and edge information are suggested. The model presented in [109] combines the perceptual notions of edge/shape information with gray level homogeneity. The authors attempt to form a unified approach by integrating the two types of information to enhance the robustness to noise and improve the poor initialization. Then, Zhu and Yuille[110] present a novel statistical and variational approach to image segmentation based on an algorithm called region competition. A more general expression of energy function is obtained, which is

$$E(C(p), \{a_i\}) = \sum_{i=1}^{M} \left\{ \frac{\mu}{2} \int_{\partial R_i} \mathrm{d}s - \right.$$
$$\left. \log P(\{I(x,y) : (x,y) \in R\}|a_i) + \lambda \right\}. \quad (3)$$

As demonstrated in [110], the snake model is a special case of (3) and can be directly deduced with fixed parameters a_i in (3). Moreover, other models, such as geodesic active region (GAR) model[111], diffusion snake[112], and polar snake[113], have improved performance in some aspects.

4.3 Discussions

Thus far, the investigation of region-based approaches is relatively limited in existing publications. More global statistics are considered in region-based approaches than in edge-based approaches. Therefore, these approaches are more robust to noise and can be adapted to the contours which are not very smooth. Specially, in the case of region-based active contour models, they can even overcome the limitation of strong dependence on initial configuration. However, the disadvantage is also obvious. That is, the approaches are invalid for contours that do not arise from region boundaries. This limitation may be overcome by the idea from [73] that open curves can be represented by circles through adding several connections between duplicated nodes. Thus, a more general region-based approach can be derived and the application of this approach will be further extended.

5 Deep-learning based approach

5.1 Existing methods

Deep convolutional neural networks (DCNNs) have re-

cently made impressive performances in various tasks such as classification, image and video detection, and segmentation[114]. The notable network (named AlexNet) is the one from Krizhevsky et al.[115], which is the first to achieve a significant improvement in large-scale image classification. Subsequently, a number of attempts have been made to improve the original architecture of Alext-Net, several of which are prevalent and worth mentioning. Their differences lie mainly in the depth and convolutional strategy. With AlexNet′s depth being 9 layers, G net[116] and GoogLeNet[117] are pushed to 16–19 layers and 22 layers, respectively. Concerning with the convolutional strategy, VGG net adopts smaller convolutional filters while GoogLeNet takes the "Inception module" to increase the width of the net. In GoogLeNet, the "Inception module" is used only at higher layers while the lower layers are kept in traditional convolutional fashion. He et al.[118] suggest a novel residual learning framework, based on which they design the named residual nets with the depth up to 152 layers. The nets are 8 deeper than the VGG nets and are substantially deeper than those used previously. The nets are successfully trained with the proposed learning frameworks, and have shown impressive performances in ILSVRC & COCO 2015 competitions. In these nets, several additional strategies are also employed, such as batch normalization[119] and dropout[120]. Inheriting the ideas of LeCun originally proposed[121], these models are typically referred to as standard DCNNs. According to the previous publications, the DCNNs that function as hierarchical feature extractors deliver strikingly better results than systems relying on carefully engineered representations.

However, despite of their great success in image classification, standard DCNNs cannot be directly applied to pixel-wise tasks, such as segmentation and detection. The repeated combination of max-pooling and striding at consecutive layers of the networks reduces significantly the spatial resolution of the resulting feature maps. In light of this, a number of effective improvements on standard DCNNs are made and new achievements have been obtained. Fully convolutional network (FCN)[122] converts the fully-connected layers of VGG into convolutional ones and attempts to harness information from multiple layers to better estimate the object boundaries. Within the same FCN framework, denser score maps are obtained by using atrous (dilated) convolutions[123–125] and the label map is refined with a fully connected CRF in [123, 124]. The networks are shown to produce high resolution segmentation. The upsampling layer employed firstly in FCN is extended to deep deconvolutional networks in [126, 127] which are composed of deconvolution, unpooling and rectified linear unit (ReLU) layers. As demonstrated, the network[126] performs significantly better than FCN even requiring more complex training and inference. The pressure can be alleviated with the architecture proposed in [127] by discarding the fully connected layers in

the VGG net. Furthermore, being fully convolutional[128] presents the feature extractor named OverFeat, based on which the integrated framework is presented for classification, localization and detection.

An orthogonal line of work on segmentation focuses on semi-supervised or weakly-supervised learning, aiming to alleviate the problem of inadequate training data by exploiting weak annotations. The methods investigate the problem of DCNNs learning for segmentation from either weakly annotated training data[129, 130] or a combination of few strongly labeled and many weakly labeled data[131, 132]. This is reasonable since weak annotations require only a fraction of efforts compared with strong annotations. Concerning with the difference in weak annotations[129, 132], use bounding box annotations while[130, 131] are based only on image-level annotations. Using these weak annotations and sometimes few strong annotations, the methods achieve results that are competitive with those of some fully supervised methods.

Girshick et al.[133] explicitly point out the differences between detection and classification. Detection requires localizing objects within an image while classification is not. To solve the detection task, a method called region convolutional neural network (RCNN) is proposed in [133]. Given an input image, the method first generates region proposals, extracts a fixed-length feature vector from each proposal using AlexNet and then classifies each region with linear SVMs. Considering the excellent performance of the method on detection, several variants are proposed to further improve the computational efficiency, including SPP net[134] and faster RCNN[135].

A popular means of introducing deep learning techniques into the contour detection problem is to replace the traditional hand-designed features by deep features. For example, N4-fields[136] extracts patch features from the pre-trained AlexNet, and then maps them to the nearest neighbor annotation from a predefined dictionary. Shen et al.[137] cluster contour patches for shape classes and then learn the deep features for contour detection by using the proposed positive-sharing loss function. The authors adopt a convolutional neural network (CNN) architecture that is simpler than the usual one. In their opinion, the contour can be represented by a local image patch with smaller size than generic objects. While the above methods in [136, 137] can be categorized as path-level feature extraction and pixel-level feature extraction[138, 139], which extracts deep feature per pixel and classifies it to edge or non-edge class. In [139], a top-down detection method is proposed, which constituted the first attempt to use higher-level object cues to predict contours. The authors insist that higher-level object-features leads to a substantial gain in contour detection accuracy.

Another series of studies could be summarized as image-level prediction[140–143], which predicts the entire contour map in an end-to-end manner. Motivated by the ar-

chitecture of the VGG net, Xie and Tu[140] propose the holistically-nested edge detection (HED) method which performs holistic image training and prediction. By adopting deep supervision strategy[144] and fully utilizing multiple side outputs, the HED method achieves a state-of-the-art performance. On this basis, reference[141] suggests a similar diverse supervision strategy that can fit to all of the intermediate layers for edge detection. In contrast to the HED method, supervision in [141] varies from coarse level to fine level as deep features become more discriminative. Liu et al.[142] propose a novel network that combines multi-scale and multi-level information from all the convolutional layers to realize image-to-image edge detection. In fact, it is the first attempt to adopt such rich convolutional features. In [7], a fully convolutional encoder-decoder network is proposed, aiming to detect higher-level object contours. The architecture takes advantage of high efficiency from FCN[122] and high resolution achieved by deconvolutional network[126], welcoming a superior performance.

However, the approach proposed in [145] is very different from the approaches mentioned above, which thus cannot be included in those classes. In this method, a variety of situations are defined and a specialized object boundary detector is trained for each of them. Given an input image, the situation is first determined based on the features extracted by AlexNet. Only those boundary detectors are applied. In addition, as similar as segmentation problem, weak-supervision techniques for contour detection are also investigated since both segmentation images and instance-wise boundaries are among the most expensive types of annotations. To relax the requirement of carefully annotated images, Khoreva et al.[146] show that bounding box annotations alone are sufficient to reach high-quality object boundaries. In addition, they propose several means of generating object boundary annotations with different levels of supervision.

5.2 Discussions

Recent CNN-based methods for contour detection demonstrate promising improvements in performance on traditional methods. Compared with the classification networks such as VGG net[115] and AlexNet[116], these methods adopt a similar architecture but different implementation strategies. Different deep learning architectures and implementation strategies are summarized in Fig. 5[140]. In particular, Figs. 5(b) and 5(c) represent multi-level and multi-scale information, respectively. In the method described in [142], both of these two types of information are involved in edge detection. This marks the first attempt to adopt such rich convolutional features in computer vision tasks. As pointed out by the authors, this can be seen as a development direction for fully connected network, like FCN[122] and HED[140]. It is also noteworthy that in several recent methods, deep network is employed only in feature extraction, but in the stage of classification traditional machine learning is applied, such as nearest neighbor search[136], SVM[137] and structured forest[137]. This strategy is firstly studied by Girshick et al.[133] and intends to further improve the prediction accuracy. Another consideration is the requirement of extensive training data. This problem has become more prominent for contour detection as labeling object boundaries is one of the most expensive types of annotations. In [146], this issue is addressed and it is shown that bounding box annotations alone are sufficient to reach high-quality object boundaries. However, researches on this topic are limited. This could be another development direction for contour detection. Methods that utilize more weak annotations such as image-level annotations[130] need further study.
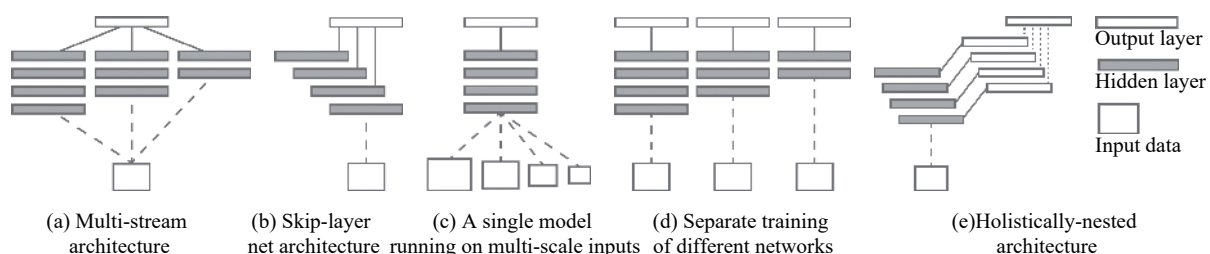

(a) Multi-stream architecture (b) Skip-layer net architecture (c) A single model running on multi-scale inputs (d) Separate training of different networks (e) Holistically-nested architecture

Fig. 5   Illustration of different multi-scale deep learning architecture configurations[140]

6 Challenging issues

As concluded from the literatures, a typical trend is to combine different kinds of features in contour detection. The first attempt is made in [4] and is extended in [49], where a large number of simple features are employed. In [142], the proposed network uses multi-level and multi-scale information simultaneously to obtain high-quality edges. However, the sufficiency and necessity of the features is not completely explored. Hence, it would be a promising research topic to analyze the interrelationship among these features and then reduce feature dimension, so as to improve the computational efficiency.

Another consideration is to introduce prior shape information about the contours, so that so as to improve contours will be detected. This is important for some spe-

cialized tasks such as tracking. The methods should have the ability to reject undesired contours and detect objects even in the presence of occlusion. A related approach is presented in [147], which aims to detect semantic contours. Moreover, the accuracy of approach in [147] is not satisfactory, which presents a particularly attractive area.

Finally, another future research direction is to analyze the relationship between semantic visual attributes and trained DCNN. On this basis, human experience can be fully utilized and the less critical features learned by DCNN can be canceled to further improve the efficiency. In addition, it is possible to control the trade-off between recall and precision, which is important for the tasks requiring high recall regardless of precision or vice versa.

7 Conclusions

In this paper, we have discussed and reviewed various approaches for contour detection that have been proposed in the last two decades. To distinguish the edge detection, the difference among the edge, the boundary and the contour is first introduced. Then, traditional contour extraction approaches are combed and analyzed in detail, which are divided into three categories. The differences and relations among those three categories are revealed. Furthermore, we also present a survey on recent theoretical progress in DCNN-based methods for contour detection. Finally, we suggest several potentially developmental issues that are valuable to be further investigated.

References

[1]   G. Papari, N. Petkov. Edge and line oriented contour detection: State of the art. *Image and Vision Computing*, vol. 29, no. 2–3, pp. 79–103, 2011. DOI: 10.1016/j.imavis.2010.08.009.

[2]   Y. S. Ming, H. D. Li, X. M. He. Contour completion without region segmentation. *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 3597–3611, 2016. DOI: 10.1109/TIP.2016.2564646.

[3]   D. Ziou, S. Tabbone. Edge detection techniques-an overview. *Pattern Recognition & Image Analysis*, vol. 8, no. 4, pp. 537–559, 1998.

[4]   D. R. Martin, C. C. Fowlkes, J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 530–549, 2004. DOI: 10.1109/TPAMI.2004.1273918.

[5]   P. Arbelaez, M. Maire, C. Fowlkes, J. Malik. Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 898–916, 2011. DOI: 10.1109/TPAMI.2010.161.

[6]   Y. LeCun, Y. Bengio, G. Hinton. Deep learning. *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. DOI: 10.1038/nature14539.

[7]   J. M. Yang, B. Price, S. Cohen, H. Lee, M. H. Yang. Object contour detection with a fully convolutional encoder-decoder network. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Las Vegas, USA, pp. 193–202, 2016. DOI: 10.1109/CVPR.2016.28.

[8]   M. S. Nixon, A. S. Aguado. *Feature Extraction & Image Processing for Computer Vision*, 3rd ed., Amsterdam, Holland: Elsevier, 2012.

[9]   J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, 1986. DOI: 10.1109/TPAMI.1986.4767851.

[10]   X. Wang. Laplacian operator-based edge detectors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 5, pp. 886–890, 2007. DOI: 10.1109/TPAMI.2007.1027.

[11]   P. Perona, J. Malik. Detecting and localizing edges composed of steps, peaks and roofs. In *Proceedings of the 3rd International Conference on Computer Vision*, IEEE, Osaka, Japan, pp. 52–57, 1990. DOI: 10.1109/ICCV.1990.139492.

[12]   Y. Lu, R. C. Jain. Reasoning about edges in scale space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 4, pp. 450–468, 1992. DOI: 10.1109/34.126806.

[13]   P. Perona, J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990. DOI: 10.1109/34.56205.

[14]   J. Malik, S. Belongile, T. Leung, J. B. Shi. Contour and texture analysis for image segmentation. *International Journal of Computer Vision*, vol. 43, no. 1, pp. 7–27, 2001. DOI: 10.1023/A:1011174803800.

[15]   N. Petkov, M. A. Westenberg. Suppression of contour perception by band-limited noise and its relation to non-classical receptive field inhibition. *Biological Cybernetics*, vol. 88, no. 3, pp. 236–246, 2003. DOI: 10.1007/s00422-002-0378-2.

[16]   D. J. Field, A. Hayes, R. F. Hess. Contour integration by the human visual system: Evidence for a local "association field". *Vision Research*, vol. 33, no. 2, pp. 173–193, 1993. DOI: 10.1016/0042-6989(93)90156-Q.

[17]   D. H. Hubel, T. N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, vol. 195, no. 1, pp. 215–243, 1968. DOI: 10.1113/jphysiol.1968.sp008455.

[18]   H. E. Jones, K. L. Grieve, W. Wang, A. M. Sillito. Surround suppression in primate V1. *Journal of Neurophysiology*, vol. 86, no. 4, pp. 2011–2028, 2001. DOI: 10.1152/jn.2001.86.4.2011.

[19]   P. Series, J. Lorenceau, Y. Fregnac. The "silent" surround of V1 receptive fields: Theory and experiments. *Journal of Physiology-Paris*, vol. 97, no. 4–6, pp. 453–474, 2003. DOI: 10.1016/j.jphysparis.2004.01.023.

[20]   J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, vol. 2, no. 7, pp. 1160–1169, 1985. DOI: 10.1364/JOSAA.2.001160.

[21]   J. P. Jones, L. A. Palmer. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, vol. 58, no. 6, pp. 1233–1258, 1987. DOI: 10.1152/jn.1987.58.6.1233.

[22] W. Chan, G. Coghill. Text analysis using local energy. *Pattern Recognition*, vol. 34, no. 12, pp. 2523–2532, 2001. DOI: 10.1016/S0031-3203(00)00155-2.

[23] R. C. Zhao, M. Wu, X. Y. Liu, B. J. Zou, F. F. Li. Orientation histogram-based center-surround interaction: An integration approach for contour detection. *Neural Computation*, vol. 29, no. 1, pp. 171–193, 2017. DOI: 10.1162/NECO_a_00911.

[24] M. P. Sceniak, D. L. Ringach, M. J. Hawken, R. Shapley. Contrast′s effect on spatial summation by macaque V1 neurons. *Nature Neuroscience*, vol. 2, no. 8, pp. 733–739, 1999. DOI: 10.1038/11197.

[25] M. P. Sceniak, M. J. Hawken, R. Shapley. Visual spatial characterization of macaque V1 neurons. *Journal of Neurophysiology*, vol. 85, no. 5, pp. 1873–1887, 2001. DOI: 10.1152/jn.2001.85.5.1873.

[26] J. R. Cavanaugh, W. Bair, J. A. Movshon. Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *Journal of Neurophysiology*, vol. 88, no. 5, pp. 2530–2546, 2002. DOI: 10.1152/jn.00692.2001.

[27] C. Grigorescu, N. Petkov, M. A. Westenberg. Contour detection based on nonclassical receptive field inhibition. *IEEE Transactions on Image Processing*, vol. 12, no. 7, pp. 729–739, 2003. DOI: 10.1109/TIP.2003.814250.

[28] C. Grigorescu, N. Petkov, M. A. Westenberg. Contour and boundary detection improved by surround suppression of texture edges. *Image and Vision Computing*, vol. 22, no. 8, pp. 609–622, 2004. DOI: 10.1016/j.imavis.2003.12.004.

[29] N. Petkov, M. A. Westenberg. Suppression of contour perception by band-limited noise and its relation to nonclassical receptive field inhibition. *Biological Cybernetics*, vol. 88, no. 3, pp. 236–246, 2003. DOI: 10.1007/s00422-002-0378-2.

[30] G. Papari, N. Petkov. An improved model for surround suppression by steerable filters and multilevel inhibition with application to contour detection. *Pattern Recognition*, vol. 44, no. 9, pp. 1999–2007, 2011. DOI: 10.1016/j.patcog.2010.08.013.

[31] G. Papari, P. Campisi, N. Petkov, A. Neri. A biologically motivated multiresolution approach to contour detection. *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Article number 071828, 2007. DOI: 10.1155/2007/71828.

[32] M. Ursino, G. E. La Cara. A model of contextual interactions and contour detection in primary visual cortex. *Neural Networks*, vol. 17, no. 5–6, pp. 719–735, 2004. DOI: 10.1016/j.neunet.2004.03.007.

[33] L. Long, Y. J. Li. Contour detection based on the property of orientation selective inhibition of non-classical receptive field. In *Proceedings of IEEE Conference on Cybernetics and Intelligent Systems*, IEEE, Chengdu, China, pp. 1002–1006, 2008. DOI: 10.1109/ICCIS.2008.4670920.

[34] C. Zeng, Y. J. Li, C. Y. Li. Center-surround interaction with adaptive inhibition: A computational model for contour detection. *NeuroImage*, vol. 55, no. 1, pp. 49–66, 2011. DOI: 10.1016/j.neuroimage.2010.11.067.

[35] K. F. Yang, S. B. Gao, C. F. Guo, C. Y. Li, Y. J. Li. Boundary detection using double-opponency and spatial sparseness constraint. *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2565–2578, 2015. DOI: 10.1109/TIP.2015.2425538.

[36] G. Azzopardi, N. Petkov. A CORF computational model of a simple cell that relies on LGN input outperforms the Gabor function model. *Biological Cybernetics*, vol. 106, no. 3, pp. 177–189, 2012. DOI: 10.1007/s00422-012-0486-6.

[37] G. Azzopardi, A. Rodriguez-Sanchez, J. Piater, N. Petkov. A push-pull CORF model of a simple cell with antiphase inhibition improves SNR and contour detection. *PLoS One*, vol. 9, no. 7, Article number e98424, 2014. DOI: 10.1371/journal.pone.0098424.

[38] K. F. Yang, C. Y. Li, Y. J. Li. Multifeature-based surround inhibition improves contour detection in natural images. *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5020–5032, 2014. DOI: 10.1109/TIP.2014.2361210.

[39] U. Polat, K. Mizobe, M. W. Pettet, T. Kasamatsu, A. M. Norcia. Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature*, vol. 391, no. 6667, pp. 580–584, 1998. DOI: 10.1038/35372.

[40] S. C. Yen, L. H. Finkel. Extraction of perceptually salient contours by striate cortical networks. *Vision Research*, vol. 38, no. 5, pp. 719–741, 1998. DOI: 10.1016/S0042-6989(97)00197-1.

[41] T. Hansen, H. Neumann. A recurrent model of contour integration in primary visual cortex. *Journal of Vision*, vol. 8, no. 8, pp. 8, 2008. DOI: 10.1167/8.8.8.

[42] Q. L. Tang, N. Sang, T. X. Zhang. Contour detection based on contextual influences. *Image and Vision Computing*, vol. 25, no. 8, pp. 1282–1290, 2007. DOI: 10.1016/j.imavis.2006.08.007.

[43] Q. L. Tang, N. Sang, T. X. Zhang. Extraction of salient contours from cluttered scenes. *Pattern Recognition*, vol. 40, no. 11, pp. 3100–3109, 2007. DOI: 10.1016/j.patcog.2007.02.009.

[44] K. F. Yang. Computational models and applications of the early stages of biological visual system, Ph. D. dissertation, University of Electronic and Technology of China, China, 2016. (in Chinese)

[45] N. Widynski, M. Mignotte. A particle filter framework for contour detection. In *Proceedings of the 12th European Conference on Computer Vision*, Springer, Florence, Italy, pp. 780–793, 2012. DOI: 10.1007/978-3-642-33718-5_56.

[46] N. Widynski, M. Mignotte. A MultiScale particle filter framework for contour detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 10, pp. 1922–1935, 2014. DOI: 10.1109/TPAMI.2014.2307856.

[47] X. F. Ren. Multi-scale improves boundary detection in natural images. In *Proceedings of the 10th European Conference on Computer Vision*, Springer, Marseille, France, pp. 533–545, 2008. DOI: 10.1007/978-3-540-88690-7_40.

[48] X. F. Ren, L. F. Bo. Discriminatively trained sparse code gradients for contour detection. In *Proceedings of the 25th International Conference on Neural Information Processing Systems*, Curran Associates Inc., Lake Tahoe, Spain, pp. 584–592, 2012.

[49] P. Dollar, Z. W. Tu, S. Belongie. Supervised learning of edges and object boundaries. In *Proceedings of IEEE Computer Society Conference on Computer Vision and*

Pattern Recognition, IEEE, New York, USA, pp. 1964–1971, 2006. DOI: 10.1109/CVPR.2006.298.

[50] J. B. Shi, J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000. DOI: 10.1109/34.868688.

[51] M. Maire, P. Arbelaez, C. Fowlkes, J. Malik. Using contours to detect and localize junctions in natural images. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Anchorage, USA, pp. 1–8, 2008. DOI: 10.1109/CVPR.2008.4587420.

[52] P. Arbelaez, J. Pont-Tuset, J. Barron, F. Marques, J. Malik. Multiscale combinatorial grouping. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Columbus, USA, pp. 328–335, 2014. DOI: 10.1109/CVPR.2014.49.

[53] J. Pont-Tuse, P. Arbelaez, J. T. Barron, F. Marques, J. Malik. Multiscale combinatorial grouping for image segmentation and object proposal generation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 1, pp. 128–140, 2017. DOI: 10.1109/TPAMI.2016.2537320.

[54] Y. Lu, R. C. Jain. Behavior of edges in scale space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 4, pp. 337–356, 1989. DOI: 10.1109/34.19032.

[55] T. Lindeberg. Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision*, vol. 30, no. 2, pp. 117–156, 1998. DOI: 10.1023/A:1008097225773.

[56] I. Kokkinos. Highly accurate boundary detection and grouping. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, San Francisco, USA, pp. 2520–2527, 2010. DOI: 10.1109/CVPR.2010.5539956.

[57] S. Wang, T. Kubota, J. M. Siskind, J. Wang. Salient closed boundary extraction with ratio contour. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 546–561, 2005. DOI: 10.1109/TPAMI.2005.84.

[58] I. J. Cox, J. M. Rehg, S. Hingorani. A Bayesian multiple-hypothesis approach to edge grouping and contour segmentation. *International Journal of Computer Vision*, vol. 11, no. 1, pp. 5–24, 1993. DOI: 10.1007/BF01420590.

[59] A. Amir, M. Lindenbaum. A generic grouping algorithm and its quantitative analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 2, pp. 168–185, 1998. DOI: 10.1109/34.659934.

[60] S. Mahamud, L. R. Williams, K. K. Thornber. Segmentation of multiple salient closed contours from real images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 4, pp. 433–444, 2003. DOI: 10.1109/TPAMI.2003.1190570.

[61] J. H. Elder, S. W. Zucker. Computing contour closure. In *Proceedings of the 4th European Conference on Computer Vision*, Springer, Cambridge, UK, pp. 399–412, 1996. DOI: 10.1007/BFb0015553.

[62] W. S. Geisler, J. S. Perry, B. J. Super, D. P. Gallogly. Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, vol. 41, no. 6, pp. 711–724, 2001. DOI: 10.1016/S0042-6989(00)00277-7.

[63] K. Thornber, L. R. Williams. Analytic solution of stochastic completion fields. *Biological Cybernetics*,

[64] L. R. Williams, D. W. Jacobs. Stochastic completion fields: A neural model of illusory contour shape and salience. *Neural Computation*, vol. 9, no. 4, pp. 837–858, 1997. DOI: 10.1162/neco.1997.9.4.837.

[65] A. Sha′shua, S. Ullman. Structural saliency: The detection of globally salient structures using a locally connected network. In *Proceedings of the 2nd International Conference on Computer Vision*, IEEE, Tampa, USA, 1988. DOI: 10.1109/CCV.1988.590008.

[66] L. Herault, R. Horaud. Figure-ground discrimination: A combinatorial optimization approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 899–914, 1993. DOI: 10.1109/34.232076.

[67] S. Sarkar, K. L. Boyer. Quantitative measures of change based on feature organization: Eigenvalues and eigenvectors. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, San Francisco, USA, pp. 478–483, 1996. DOI: 10.1109/CVPR.1996.517115.

[68] G. Guy, G. Medioni. Inferring global pereeptual contours from local features. *International Journal of Computer Vision*, vol. 20, no. 1–2, pp. 113–133, 1996. DOI: 10.1007/BF00144119.

[69] L. R. Williams. A comparison of measures for detecting natural shapes in cluttered backgrounds. *International Journal of Computer Vision*, vol. 34, no. 2–3, pp. 81–96, 1999. DOI: 10.1023/A:1008187804026.

[70] I. H. Jermyn, H. Ishikawa. Globally optimal regions and boundaries as minimum ratio weight cycles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1075–1088, 2001. DOI: 10.1109/34.954599.

[71] T. Schoenemann, S. Masnou, D. Cremers. The elastic ratio: Introducing curvature into ratio-based image segmentation. *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2565–2581, 2011. DOI: 10.1109/TIP.2011.2118225.

[72] J. S. Stahl, S. Wang. Globally optimal grouping for symmetric closed boundaries by combining boundary and region information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 395–411, 2008. DOI: 10.1109/TPAMI.2007.1186.

[73] Q. H. Zhu, G. Song, J. B. Shi. Untangling cycles for contour grouping. In *Proceedings of the 11th IEEE Conference on Computer Vision*, IEEE, Rio de Janeiro, Brazil, pp. 1–8, 2007. DOI: 10.1109/ICCV.2007.4408929.

[74] M. Meila, J. B. Shi. Learning segmentation by random walks. In *Proceedings of the 14th Annual Neural Information Processing Systems*, NIPS, Denver, USA, pp. 839–873, 2001.

[75] I. Kokkinos. Highly accurate boundary detection and grouping. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, San Francisco, USA, pp. 2520–2527, 2010.(本条文献与第56条文献重复请核对) DOI: 10.1109/CVPR.2010.5539956.

[76] B. Fischer, J. M. Buhmann. Path-based clustering for grouping of smooth curves and texture segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 4, pp. 513–518, 2003. DOI: 10.1109/TPAMI.2003.1190577.

[77] P. Felzenszwalb, D. McAllester. A min-cover approach

for finding salient curves. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshop*, IEEE, New York, USA, pp. 185, 2006. DOI: 10.1109/CVPRW.2006.18.

[78] D. W. Jacobs. Robust and efficient detection of salient convex groups. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 1, pp. 23–37, 1996. DOI: 10.1109/34.476008.

[79] Y. S. Ming, H. D. Li, X. M. He. Connected contours: A new contour completion model that respects the closure effect. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Providence, USA, pp. 892–836, 2012. DOI: 10.1109/CVPR.2012.6247755.

[80] X. F. Ren, C. C. Fowlkes, J. Malik. Learning probabilistic models for contour completion in natural images. *International Journal of Computer Vision*, vol. 77, no. 1–3, pp. 47–63, 2008. DOI: 10.1007/s11263-007-0092-6.

[81] M. Kass, A. Witkin, D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988. DOI: 10.1007/BF00133 570.

[82] V. Caselles, R. Kimmel, G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, vol. 22, no. 1, pp. 61–79, 1997. DOI: 10.1023/A:1007979827043.

[83] V. Caselles, R. Kimmel, G. Sapiro. Geodesic active contours. In *Proceedings of International Conference on Computer Vision*, IEEE, Cambridge, USA, pp. 694–699, 1995. DOI: 10.1109/ICCV.1995.466871.

[84] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, A. Yezzi. Gradient flows and geometric active contour models. In *Proceedings of the 5th IEEE International Conference on Computer vision*, IEEE, Cambridge, USA, pp. 810–815, 1995. DOI: 10.1109/ICCV.1995.466855.

[85] A. Blake, M. Isard. *Active Contours: The Application of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion*, Berlin, Germany: Springer, pp. 49–103, 2012. (请核对年份信息)

[86] L. D. Cohen. On active contour models and balloons. *CVGIP: Image Understanding*, vol. 53, no. 2, pp. 211–218, 1991. DOI: 10.1016/1049-9660(91)90028-N.

[87] L. D. Cohen, L. Cohen. Finite-element methods for active contour models and balloons for 2D and 3D images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1131–1147, 1993. DOI: 10.1109/34.244675.

[88] C. Y. Xu, J. L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 359–369, 1998. DOI: 10.1109/83.661186.

[89] O. Amadieu, E. Debreuve, M. Barlaud, G. Aubert. Inward and outward curve evolution using level set method. In *Proceedings of International Conference on Image Processing*, IEEE, Kobe, Japan, pp. 188–192, 1999. DOI: 10.1109/ICIP.1999.817097.

[90] M. E. Leventon, W. E. L. Grimson, O. Faugeras. Statistical shape influence in geodesic active contours. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Hilton Head Island, USA, pp. 316–323, 2000. DOI: 10.1109/CVPR.2000.855835.

[91] A. Dumitras, A. N. Venetsanopoulos. Angular map-driven snakes with application to object shape description in color images. *IEEE Transactions on Image Processing*, vol. 10, no. 12, pp. 1851–1859, 2001. DOI: 10.1109/83. 974570.

[92] L. H. Staib, J. S. Duncan. Boundary finding with parametrically deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 11, pp. 1061–1075, 1992. DOI: 10.1109/34.166621.

[93] S. R. Gunn, M. S. Nixon. A robust snake implementation: A dual active contour. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 1, pp. 63–68, 1997. DOI: 10.1109/34.566812.

[94] B. Al-Diri, A. Hunter, D. Steel. An active contour model for segmenting and measuring retinal vessels. *IEEE Transactions on Medical imaging*, vol. 28, no. 9, pp. 1488–1497, 2009. DOI: 10.1109/TMI.2009.2017941.

[95] J. Park, J. M. Keller. Snakes on the watershed. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1201–1205, 2001. DOI: 10.1109/34.954609.

[96] H. W. Park, T. Schoepflin, Y. Kim. Active contour model with gradient directional information: Directional snake. *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 2, pp. 252–256, 2001. DOI: 10.1109/76.905991.

[97] J. S. Suri, K. C. Liu, S. Singh, S. N. Laxminarayan, X. L. Zeng, L. Reden. Shape recovery algorithms using level sets in 2D/3D medical imagery: A state-of-the-art review. *IEEE Transactions on Information Technology in Biomedicine*, vol. 6, no. 1, pp. 8–28, 2002. DOI: 10.1109/4233.992158.

[98] W. J. Zhou, Z. X. Fei, H. S. Hu, L. Liu, J. N. Li, P. J. Smith. Real-time object subspace searching based on discrete searching paths and local energy. *International Journal of Automation and Computing*, vol. 13, no. 2, pp. 99–107, 2016. DOI: 10.1007/s11633-015-0946-2.

[99] N. R. Pal, S. K. Pal. A review on image segmentation techniques. *Pattern Recognition*, vol. 26, no. 9, pp. 1277–1294, 1993. DOI: 10.1016/0031-3203(93)90135-J.

[100] H. D. Cheng, X. H. Jiang, Y. Sun, J. L. Wang. Color image segmentation: Advances and prospects. *Pattern Recognition*, vol. 34, no. 12, pp. 2259–2281, 2001. DOI: 10.1016/S0031-3203(00)00149-7.

[101] R. C. Zhao, Y. D. Ma. A region segmentation method for region-oriented image compression. *Neurocomputing*, vol. 85, pp. 45–52, 2012. DOI: 10.1016/j.neucom.2012.01.007.

[102] P. Arbelaez. Boundary extraction in natural images using ultrametric contour maps. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, New York, USA, pp. 182–182, 2006. DOI: 10.1109/CVPRW.2006.48.

[103] P. Arbelaez, M. Maire, C. Fowlkes, J. Malik. From contours to regions: An empirical evaluation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Miami, USA, pp. 2294–2301, 2009. DOI: 10.1109/CVPR.2009.5206707.

[104] Y. S. Ming, H. D. Li, X. M. He. Winding number constrained contour detection. *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 68–79, 2015. DOI: 10.1109/TIP.2014.2372636.

[105] Y. S. Ming, H. D. Li, X. M. He. Winding number for region-boundary consistent salient contour extraction. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Portland, USA, pp. 2818–2825,

2013. DOI: 10.1109/CVPR.2013.363.

[106] L. D. Cohen, E. Bardnient, N. Ayache. Surface reconstruction using active contour models. IA, IRA, Sophia-Antipolis, France, 1993.

[107] T. F. Chan, L. A. Vese. Active contours without edges. IEEE Transactions on Image Processing, vol. 10, no. 2, pp. 266–277, 2001. DOI: 10.1109/83.902291.

[108] L. A. Vese, T. F. Chan. A multiphase level set framework for image segmentation using the Mumford and Shah model. International Journal of Computer Vision, vol. 53, no. 3, pp. 271–293, 2002. DOI: 10.1023/A:1020874308076.

[109] A. Chakraborty, L. H. Staib, J. S. Duncan. Deformable boundary finding in medical images by integrating gradient and region information. IEEE Transactions on Medical Imaging, vol. 15, no. 6, pp. 859–870, 1996. DOI: 10.1109/42.544503.

[110] S. C. Zhu, A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, no. 9, pp. 884–900, 1996. DOI: 10.1109/34.537343.

[111] N. Paragios, R. Deriche. Geodesic active regions and level set methods for supervised texture segmentation. International Journal of Computer Vision, vol. 46, no. 3, pp. 223–247, 2002. DOI: 10.1023/A:1014080923068.

[112] D. Cremers, F. Tischhuser, J. Weickert, C. Schnorr. Diffusion snakes: Introducing statistical shape knowledge into the Mumford-Shah functional. International Journal of Computer Vision, vol. 50, no. 3, pp. 295–313, 2002. DOI: 10.1023/A:1020826424915.

[113] C. Collewet. Polar snakes: A fast and robust parametric active contour model. In Proceedings of the 16th IEEE International Conference on Image Processing, IEEE, Cairo, Egypt, pp. 3013–3016, 2009. DOI: 10.1109/ICIP.2009.5414444.

[114] B. Zhao, J. S. Feng, X. Wu, S. C. Yan. A survey on deep learning-based fine-grained object classification and semantic segmentation. International Journal of Automation and Computing, vol. 14, no. 2, pp. 119–135, 2017. DOI: 10.1007/s11633-017-1053-3.

[115] A. Krizhevsky, I. Sutskever, G. E. Hinton. ImageNet classification with deep convolutional neural networks. In Proceedings of Advances in Neural Information Processing Systems, NIPS, Lake Tahoe, Spain, pp. 1097–1105, 2012. DOI: 10.1145/3065386.

[116] K. Simonyan, A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv: 1409.1556, 2014.

[117] C. Szegedy, W. Liu, Y. Q. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich. Going deeper with convolutions. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Boston, USA, pp. 1–9, 2015. DOI: 10.1109/CVPR.2015.7298594.

[118] K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun. Deep residual learning for image recognition. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Las Vegas, USA, pp. 770–778, 2016. DOI: 10.1109/CVPR.2016.90.

[119] S. Ioffe, C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on Machine Learning, IEEE, Lille, France, pp. 448–456, 2015.

[120] E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R. R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv: 1207.0580, 2012.

[121] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel. Backpropagation applied to handwritten zip code recognition. Neural Computation, vol. 1, no. 4, pp. 541–551, 1989. DOI: 10.1162/neco.1989.1.4.541.

[122] E. Shelhamer, J. Long, T. Darrell. Fully convolutional networks for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 4, pp. 640–651, 2017. DOI: 10.1109/TPAMI.2016.2572683.

[123] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected CRFs. arXiv preprint arXiv: 1412.7062, 2014.

[124] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille. DeepLab: Semantic image segmentation with deep convolutional nets, Atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 834–848, 2018. DOI: 10.1109/TPAMI.2017.2699184.

[125] F. Yu, V. Koltun. Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv: 1511.07122, 2015.

[126] H. Noh, S. Hong, B. Han. Learning deconvolution network for semantic segmentation. In Proceedings of IEEE International Conference on Computer Vision, IEEE, Santiago, Chile, pp. 1520–1528, 2015. DOI: 10.1109/ICCV.2015.178.

[127] V. Badrinarayanan, A. Kendall, R. Cipolla. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv preprint arXiv: 1511.00561, 2015.

[128] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, Y. LeCun. OverFeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv: 1312.6229, 2013.

[129] J. F. Dai, K. M. He, J. Sun. BoxSup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In Proceedings of IEEE Conference on Computer Vision, IEEE, Santiago, Chile, pp. 1635–1643, 2015. DOI: 10.1109/ICCV.2015.191.

[130] P. O. Pinheiro, R. Collobert. Weakly supervised semantic segmentation with convolutional networks. arXiv preprint arXiv: 1411.6228, 2014. (请核对修改是否正确)

[131] S. Hong, H. Noh, B. Han. Decoupled deep neural network for semi-supervised semantic segmentation. In Proceedings of Advances in Neural Information Processing Systems, NIPS, Montreal, Canada, pp. 1495–1503, 2015.

[132] G. Papandreou, L. C. Chen, K. P. Murphy, A. L. Yuille. Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. In Proceedings of IEEE International Conference on Computer Vision, IEEE, Santiago, Chile, pp. 1742–1750, 2015. DOI: 10.1109/ICCV.2015.203.

[133] R. Girshick, J. Donahue, T. Darrell, J. Malik. Rich feature hierarchies for accurate object detection and semant-

ic segmentation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Columbus, USA, pp. 580–587, 2014. DOI: 10.1109/CVPR.2014.81.

[134] K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. In *Proceedings of the 13th European Conference on Computer Vision*, Zurich, Switzerland, pp. 346–361, 2014. DOI: 10.1007/978-3-319-10578-9_23.

[135] S. Q. Ren, K. M. He, R. Girshick, J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Proceedings of Advances in Neural Information Processing*, NIPS, Montreal, Canada, pp. 91–99, 2015.

[136] Y. Ganin, V. Lempitsky. *N*.4-fields: Neural network nearest neighbor fields for image transforms. In *Proceedings of the 12th Asian Conference on Computer Vision*, Springer, Singapore, pp. 536–541, 2014. DOI: 10.1007/978-3-319-16808-1_36.

[137] W. Shen, X. G. Wang, Y. Wang, X. Bai, Z. J. Zhang. DeepContour: A deep convolutional feature learned by positive-sharing loss for contour detection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Boston, USA, pp. 3982–3991, 2015. DOI: 10.1109/CVPR.2015.7299024.

[138] J. J. Hwang, T. L. Liu. Pixel-wise deep learning for contour detection. arXiv preprint arXiv: 1504.01989, 2015.

[139] G. Bertasius, J. B. Shi, L. Torresani. DeepEdge: A multi-scale bifurcated deep network for top-down contour detection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Boston, USA, pp. 4380–4389, 2015. DOI: 10.1109/CVPR.2015.7299067.

[140] S. Xie, Z. W. Tu. Holistically-nested edge detection. In *Proceedings of IEEE Conference on Computer Vision*, IEEE, Santiago, Chile, pp. 1395–1403, 2015. DOI: 10.1109/ICCV.2015.164.

[141] Y. Liu, M. S. Lew. Learning relaxed deep supervision for better edge detection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Las Vegas, USA, pp. 231–240, 2016. DOI: 10.1109/CVPR.2016.32.

[142] Y. Liu, M. M. Cheng, X. W. Hu, K. Wang, X. Bai. Richer convolutional features for edge detection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Honolulu, USA, 2017. DOI: 10.1109/CVPR.2017.622.

[143] K. K. Maninis, J. Pont-Tuset, P. Arbelaez, L. Van Gool. Convolutional oriented boundaries. In *Proceedings of the 14th European Conference on Computer Vision*, Springer, Amsterdam, The Netherlands, pp. 580–596, 2016. DOI: 10.1007/978-3-319-46448-0_35.

[144] C. Y. Lee, S. Xie, P. W. Gallagher, Z. Y. Zhang, Z. W. Tu. Deeply-supervised nets. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics*, JMLR, San Diego, USA, pp. 562–570, 2015.

[145] J. R. R. Uijlings, V. Ferrari. Situational object boundary detection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Boston, USA, pp. 4712–4721, 2015. DOI: 10.1109/CVPR.2015.7299103.

[146] A. Khoreva, R. Benenson, M. Omran, M. Hein, B. Schiele. Weakly supervised object boundaries. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Las Vegas, USA, pp. 183–192, 2016. DOI: 10.1109/CVPR.2016.27.

[147] B. Hariharan, P. Arbelaez, L. Bourdev. Semantic contours from inverse detectors. In *Proceedings of IEEE Conference on Computer Vision*, IEEE, Barcelona, Spain, pp. 991–998, 2011. DOI: 10.1109/CVPR.2016.27.

Xin-Yi Gong received the B. Sc. degree in automation from Tsinghua University, China, in 2014. He is currently a Ph.D degree candidate in control science and engineering at Institute of Automation, Chinese Academy of Sciences, University of Chinese Academy of Sciences, China.

His research interests include computer vision, image processing, pattern recognition and machine learning.

E-mail: gongxinyi2014@ia.ac.cn

ORCID iD: 0000-0002-6515-2836
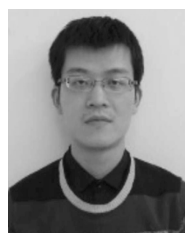
Hu Su received the B. Sc. and M. Sc. degrees from Shandong University, China, in 2007 and 2010, respectively, and the Ph.D. degree from the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences (IACAS) China in 2013. Since 2013, he has been with the IACAS, China, where he is currently an associate researcher in the Research Center of Precision Sensing and Control, China.

His research interests include intelligent control and optimization.

E-mail: hu.su@ia.ac.cn

De Xu received the B. Sc. and M. Sc. degrees in control science and engineering from Shandong University of Technology, China in 1985 and 1990, respectively, and the Ph. D. degree in control science and engineering from Zhejiang University, China in 2001. Since 2001, he has been with IACAS, China. He is currently a professor in the Research Center of Precision Sensing and Control, IACAS, China.

His research interests include robotics and automation, especially the control of robots such as visual control and intelligent control.

E-mail: de.xu@ia.ac.cn

ORCID iD: 0000-0002-7221-1654

Zheng-Tao Zhang received the B. Sc. degree in control science and engineering from China University of Petroleum, China in 2004, the M. Sc. degree in control science and engineering from Beijing Institute of Technology, China in 2007, and the Ph. D. degree in control science and engineering from IACAS, China in 2010. He is a professor in the Research Center of Precision Sensing and Control, IACAS, China.

His research interests include visual measurement, micro-assembly and automation.

E-mail: zhengtao.zhang@ia.ac.cn
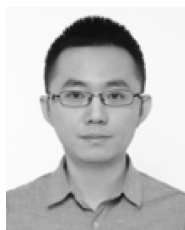
Fei Shen received the B. Sc. and M. Sc. degrees from Xidian University, China, and Beijing Institute of Technology, China, in 2007 and 2009, respectively, and the Ph. D. degree in control science and engineering from IACAS, China in 2012. He is currently an associate professor in the Research Center of Precision Sensing and Control, IACAS, China.

His research interests include robot control, robot vision control and micro-assembly.

E-mail: fei.shen@ia.ac.cn

Hua-Bin Yang received the B. Sc. degree in mechanical manufacture and automation from Shandong University of Technology, China in 2010, and the Ph. D. degree in mechanical manufacture and automation from CIOMPCAS, China in 2015. He is a research assistant professor in the Research Center of Precision Sensing and Control, IACAS, China.

His research interests include optics and precision mechanics, visual measurement and automation.

E-mail: huabin.yang@ia.ac.cn (Corresponding author)

ORCID iD: 0000-0001-9955-0560