Image Segmentation with Boundary-to-Pixel Direction and Magnitude Based on Watershed and Attention Mechanism

Hongyang Xu
Wuhan University of Science and Technology
Yuanxiu Xing (✉ yuanxiu@126.com)
Wuhan University of Science and Technology
Wenbo Wang
Wuhan University of Science and Technology

Research Article

Keywords: Image segmentation, Deep learning, Boundary-to-pixel direction and magnitude, Watershed, Attention mechanism

Posted Date: March 23rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1478176/v1

Image Segmentation with Boundary-to-Pixel Direction and Magnitude Based on Watershed and Attention Mechanism

Hongyang Xu[1,2], Yuanxiu Xing[1,2], Wenbo Wang[1,2]

Abstract

An improved image segmentation algorithm with boundary-to-pixel direction and magnitude (IS-BPDM) is proposed to deal with small regions segmentation while keeping the accuracy of edge segmentation. First, we develop a BPDM network embedded with watershed and attention module and use an adaptive loss function to achieve each pixel's robust and accurate BPDM which is defined as a two-dimensional vector, including direction and magnitude, and pointing from its nearest boundary pixel to itself. Then, we use the leaned BPDMs to obtain the refined initial segmented regions by considering the pixels near boundary have shorter magnitude and near root pixels have longer magnitude, meanwhile adjacent pixels in different regions or nearby pixels on both sides of root pixel in same region have opposite directions and nearby pixels in same region have similar directions. Last, we utilize a fast grouping method according to direction similarity to combine these initial segmented regions into final segmentation. The experimental results show that compared with the state-of-art methods in image segmentation, the IS-BPDM approach proposed in this paper achieves better segmentation accuracy and high computational efficiency, and outperforms in small regions segmentation on public datasets.

Keywords: Image segmentation · Deep learning · Boundary-to-pixel direction and magnitude · Watershed · Attention mechanism

1 Introduction

Image segmentation aims to divide an image into non-overlapping regions, and pixels in each region have their own unique perceptual appearance, e.g., color, texture, intensity. It is the basis of target detection and image classification, and has become a key step in artificial intelligence applications [1]. However, there are still great challenges in small object segmentation, accurate edge segmentation and efficiency.

Many traditional image segmentation tasks are unsupervised learning by using region [2], threshold [3–5], boundary [6], graph theory [7], energy functional [8] and so on. Though these methods have been widely used in simple structure images segmentation, often insufficient priori properties knowledge easily lead to a dissatisfied performance in dealing with the weak boundaries on natural images [9–11]. In addition, thanks to have a large cost in converting the contour into segmentation, the difficulty of implementing these methods is analogous to building a single-span bridge across a wide river.

With the development of artificial neural networks, some state-of-the-art image segmentation techniques [12–17] based on deep learning are mainly end-to-end approaches and have a witnessed significant progress in both accuracy and computational efficiency. The milestone approach is the fully connected convolutional neural network (FCN) [16] which adapts contemporary classification networks such as VGG and uses skip connection architecture, followed by an up sampled deconvolution network to accomplish semantic segmentation. With further development, novel

✉ Yuanxiu Xing
yuanxiu@126.com

Hongyang Xu
1291798923@qq.com

Wenbo Wang
wangwenbo@wust.edu.cn

[1] College of Science, Wuhan University of Science and Technology, Wuhan 430081, China

[2] Hubei Province Key Laboratory of Systems Science in Metallurgical Process, Wuhan 430081, China

segmentation approaches, such as DeepLab [17] based on altrous convolutions was proposed to handle the problem of segmenting objects at multiple scales. In order to make adequate use of the semantic context information of image scene, [18, 19] proposed a conditional generation adversarial network (cGAN) to solve the general pixel-to-pixel mapping problem, and automatically learned to segment the image accurately. [20] trained the network end-to-end, pixel-to-pixel on semantic segmentation to reduce parameter redundancy and time cost. However, most of them failed to identify weak boundary and some small objects, and then some works considered integrating traditional methods into deep network to solve the limitations.

One excellent way addressing the limitations of traditional methods and aforementioned deep learning approaches is to use boundary-to-pixel direction (BPD) [21–23] of each pixel to improve the segmentation performance. BPD is learned and is used to represent the relative position between each pixel and its nearest boundary pixel. The better performing method [21] combined the pixels with similar BPDs according to the given thresholds to form super-BPDs, so as to ensure that all pixels in the same super-BPD have robust similar directional characteristics.

Although the super-BPD [21] usually achieves a pleasant trade-off between the accuracy and efficiency on image segmentation, there are still some challenges. On the one hand, the learned BPDs are not accurate enough for weak edges and small regions because of dramatic changes of direction. On the other hand, the BPD only considers the direction of the pixel and ignores the important magnitude which represents the distance from the pixel to boundary. These two drawbacks lead to poor segmentation performance on small regions segmentation and overlapping targets. In fact, watershed has a good effect on processing overlap regions and the weak boundaries, and the attention module can make more attention to small and weak edge characteristics. Therefore, we propose a BPDM network embedded with watershed and attention module and use an adaptive loss function to learn each pixel's robust and accurate BPDM. On this basis, we use the direction similarity and magnitude of the learned BPDMs to achieve the final segmentation.

To conclude, our contributions are in these aspects:

•Proposing a novel BPDM network and loss function to obtain robust and accuracy BPDMs, which can effectively improve the accuracy of BPDMs on small regions and weak edges.

•Improving the segmentation algorithm by using the priori properties of BPDMs to refine boundary pixels and root pixels, which can lead to a pleasant image segmentation result.

•The experimental results evaluating on three datasets demonstrate that the presented segmentation approach achieves competitive performances against some state-of-the-art methods.

The rest of the paper is organized as: The related techniques are briefly described in Sect.2. Details of our BPDM learning approach is proposed in Sect.3. Image segmentation with BPDMs is introduced in Sect.4. Datasets, implementation details and experimental results are displayed in Sect.5. Then the conclusion is indicated in Sect.6.

2 Related Work

We shortly review some works on image segmentation tasks leveraging watershed algorithm, attention module and direction information.

Watershed algorithm.Watershed algorithm [24–26] regards the image as a topological terrain which is divided into catchment basin (i.e. catchment area) and watershed (i.e. dam). They are boundary segmentation methods, which achieve image segmentation by using the extracted object contour features. In many segmentation scenes, the overlap between multiple objects in an image leads to the wrong merging of smaller objects and larger regions, which is a challenge for image segmentation task. Watershed algorithm has satisfactory performance for weak edges, overlap and small regions segmentation. But only using watershed often leads to over-segmentation, and cannot merge the result pieces into one component and produce incorrect semantic segmentation when segment overlap objects. [24] used two-phase super-pixel segmentation method based on the watershed transformation with global and local boundary marching, and produced superior accuracy and computing time. Mutex watershed algorithm [25] learned local attractive and repulsive edges, followed by an improved maximum spanning tree to achieve well image segmentation. The marker watershed algorithm [26]combined watershed and end-to-end CNNs to solve the problem of complex processing steps in most examples, and improved the segmentation performance. The Otsu algorithm [27] used watershed transform to isolate cluster nuclei from each other. In this paper, watershed algorithm module is used to preprocess the original image, so that the weak edge contour of the objects can be emphasized to obtain the accurate BPDMs.

Attention module: The attention mechanism originates from imitating human visual perception and plays a vital role in the sensory system [28, 29]. The sensory system can use the focusing function to focus on some local scenes, transfer limited visual attention to the local areas of interest, and selectively capture more important visual structure information [29]. With the rise of CNNs, most of works have proved that adding attention mechanism to the CNNs structure can improve the feature expression ability of the network [30, 31]. Such as the SE module proposed in [30] introduced the attention mechanism only on the channel. CBAM proposed in [31] considered the attention mechanism from the two dimensions of channel and space. CBAM and SE modules can be embedded in mainstream network, which can not only improve the ability of model feature extraction, but also control the amount of computation. Based on these, this paper adds the CBAM to the BPDM network to learn robust BPDMs.

BPD Learning: Inspired by the algorithms of computing component trees [22, 23], BPD [21] provided direction information for each pixel and was effective informative for super-pixels [5]. It was convenient for the subsequent grouping and merging of pixels according to direction similarity that nearby pixels from different regions have opposite directions and adjacent pixels in the same region have similar directions. In [21], BPDs were learned based on FCN structure in which added ASPP layer [17] to enlarges the receptive field in down-sampling process, and then were partitioned into super-BPDs by using the robust direction similarity. Although super-BPD can separate nearby regions with weak boundaries, the segmentation on small regions is not very well due to the learned BPDs around small regions are not very accurate.

For solving this issue, a novel BPDM network is proposed, in which watershed module and CBAM module are added into FCN to learn robust and accurate BPDM of each pixel. Then, besides the direction similarity, magnitude of learned BPDMs are also used to effectively produce root pixels and initial segmentations, followed by using region adjacency graphs (RAG) partition algorithm to accomplish the final image segmentation. The proposed approach can effectively improve the segmentation accuracy of edge and small regions.

3 BPDM Learning

3.1 BPDM Definition

For each pixel p in the image, we search its nearest boundary pixel B_p, and BPDM of p is given by:

$$DM_p = \overrightarrow{B_p p} \tag{1}$$

where DM_p is a two-dimensional direction vector pointing from B_p to p.

The DM_p provides cues about direction and magnitude. The direction is used to calculate the similarity between p and other pixels, and the magnitude is used to determine whether p is a boundary pixel or root pixel.

3.2 Architecture of BPDM Network

The quality of BPDMs directly affects the performance of subsequent image segmentation. As shown in Fig.1, the proposed BPDM network includes WA feature extraction module and multi-scale feature fusion module, and learns accurate BPDM of each pixel.

3.2.1 WA Feature Extraction

Considering that many image segmentation tasks do not have enough accuracy for small regions segmentation, watershed module and attention mechanism module are embedded into WA feature extraction module to remedy this issue. As shown in Fig.1, in the WA feature extraction module, mathematical morphological transformation is used to mark the foreground and background of the image to obtain the marked input image. Then watershed module is used to extract contour features of objects to realize rough image segmentation. Thanks to the complexity of image information, the attention mechanism is also added in the down-sampling network to assign different weight information to pixel features, so that the weak boundaries and boundaries of small regions can be focused on.

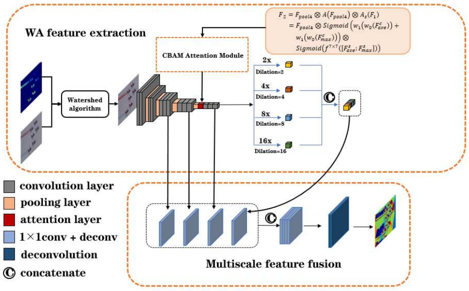

WA feature extraction
CBAM Attention Module
$F_2 = F_{pool4} \otimes A(F_{pool4}) \otimes A_s(F_1)$
$= F_{pool4} \otimes Sigmoid\left(w_1(w_0(F^c_{ave})) + w_1(w_0(F^c_{max}))\right) \otimes$
$Sigmoid(f^{7\times7}([F^s_{ave}; F^s_{max}]))$
Watershed algorithm
2x
Dilation=2
4x
Dilation=4
8x
Dilation=8
16x
Dilation=16
©
convolution layer
pooling layer
attention layer
1×1conv + deconv
deconvolution
© concatenate
Multiscale feature fusion

Fig. 1 BPDM network architecture

In forward propagation, BPDM network uses five group convolution layers and four maximum pooling layers, and embeds attention mechanism module behind the fourth pooling layer to extract the attention feature map which is used as the input of the fifth convolution layer. The feature images output by the last convolution layer are performed 2, 4, 8 and 16 times dilation respectively in ASPP layer and concatenated as an output of WA feature extraction module.

Fig.2 illustrates the details of the CBMA module which extracts features from both channel and space dimensions. This module can be integrated into any CNNs architectures seamlessly with negligible overheads and is end-to-end trainable along with base CNNs. The intermediate feature map output F_{pool4} by the fourth pooling layer is used as the input of the channel attention module, and the outputs feature map F_1 of channel attention module is used as the input feature map of the space attention module, and the final feature map is F_2. The whole process of CBAM is as follows:

$$F_1 = A_c(F_{pool4}) \otimes F_{pool4} \tag{2}$$

$$F_2 = A_s(F_1) \otimes F_1 \tag{3}$$

where $\otimes$ denotes element-wise multiplication. $A_c(\bullet)$ and $A_s(\bullet)$ are operators of channel attention and space attention respectively.

In short, the channel attention and the spatial attention are computed respectively as:

$$A_c(F_{pool4}) = \sigma(MLP(AvgPool(F_{pool4})) + MLP \\ (MaxPool(F_{pool4}))) \tag{4}$$

$$A_s(F_1) = \sigma(f^{7\times 7}([AvgPool(F_1);\ MaxPool(F_1)])) \tag{5}$$

where MLP denotes multi-layer perceptron with one hidden layer, $AvgPool$ is mean pooling layer and $MaxPool$ is maximum pooling layer. σ denotes the sigmoid function. $f^{7\times 7}$ is 7×7 convolution kernel is used for feature fusion.

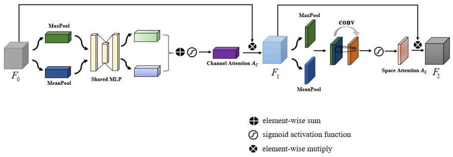


Fig. 2 CBMA: channel and space attention module

3.2.2 Multiscale Feature Fusion

As shown in Fig.1, in multi-scale feature fusion module, 1×1 convolution and deconvolution are applied into $conv3, conv4, conv5$ and the output of WA feature extraction module, followed by a skip connection of these output features. Finally, three consecutive 1×1 deconvolutions are used on the fuse feature maps to achieve the BPDMs prediction. The whole process of multiscale feature fusion is as follows:

$$\begin{aligned} M_b =& [S^{conv3}(ReLU[f^{1\times 1}(F_{conv3})];\ ReLU[f^{1\times 1} \\ &(F_{conv4})];\ ReLU[f^{1\times 1}(F_{conv5})];\ ReLU[f^{1\times 1} \\ &(F_{WA})])]_b \end{aligned} \tag{6}$$

where M_b denotes the four feature maps series, in which the maps are resized to the size of $conv3$ and are bilinear up sampled. $[\bullet]_b$ is skip connection operator. $ReLU$ is activation function. S^{conv3} represents the operation of resizing the feature map to the size of the third convolution layer. $f^{1\times 1}$ is defined as convolution of 1×1. $F_{conv3}, F_{conv4}, F_{conv5}$ and F_{WA} represent the output feature map of $conv3, conv4, conv5$ and WA feature extraction module.

Then perform the following operation to obtain BPDMs of all pixels in the image:

$$DMs = \widetilde{f}^{3\times 1\times 1}(M_b) \tag{7}$$

where $\widetilde{f}^{3\times 1\times 1}$ denotes three 1×1 deconvolution operations.

3.3 Adaptive Loss Function

The magnitude loss and the direction loss are considered for BPDMs learning. The loss function for BPDMs leaning is defined as following:

$$\begin{aligned} L &= \sum_{p\in\Omega} w(p)\,(L_m + L_d) \\ L_m &= \beta(p)\|DM_p - \hat{DM_p}\|^2 \\ L_d &= \alpha\|cos^{-1} < DM_p, \hat{DM_p} >\|^2 \end{aligned} \tag{8}$$

where $w(p) = 1/|GT_p|^n$,$n > 0$ is the adaptive weight of pixel p. $|GT_p|$ is the size of ground truth segment containing p. The larger of n, the more importance of the small regions. L_m and L_d are the loss of magnitude and direction respectively. DM_p and $\hat{DM_p}$ represent the ground truth BPDM and learned BPDM of p respectively. $\|\bullet\|^2$ is L_2 norm. $\beta(p) = 1/|D_p|$ is used to normalize the magnitude loss. α is the hyperparameter to trade off between the direction loss and magnitude loss, and generally is set to 1.

Fig.3 demonstrates the heatmap visualization of the learned results for L_2 magnitude and direction of each pixel. super-BPD [21], mainly focus on the boundary-to-pixel direction of each pixel training, and the learned magnitude of each pixel is around 1. For super-BPD, small objects with coordinate near 100 on the x-axis are not recognized. After adding watershed module and attention mechanism module into super-BPD network respectively, the results on small segmentation regions have been improved. Based on these work, the proposed BPDM network is constructed by combining watershed and attention mechanism module, which retains the smoothness of pixel direction, and the prediction effect on fine objects has been improved to a certain extent. It can be observed that the prediction results on L_2 magnitude and direction of IS-BPDM is more refine on small regions.

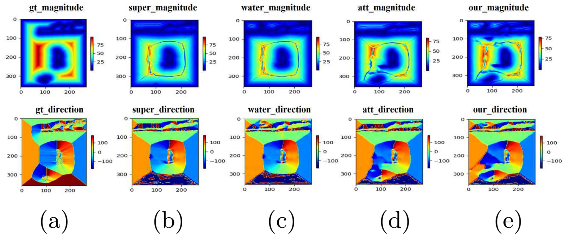


Fig. 3 Intermediate details of prediction. (a) the real intermediate images from label. (b) the intermediate images from the prediction of super-BPD [21]. (c) The intermediate images adding the watershed algorithm based on super-BPD. (d) The intermediate images adding the attention mechanism based on super-BPD. (e) The intermediate images of the proposed method which combines the watershed and the attention mechanism

4 Image Segmentation Based on BPDMs

Initial segmentation. Inspired by the algorithms of computing parent image and root pixels of each region [21], the parent image $\mathcal{P}$ and root pixels $\mathcal{R}$ are optimized according to the directions and magnitudes of learned BPDMs, as depicted in Algo.1. Initially, the parent of each pixel p is set to itself and the root pixel set $\mathcal{R}$ is empty. Then we calculate the direction between the pixel p and the neighbor pixel n_p, and compare their included angle $cos^{-1}\langle \hat{DM}_p, \hat{DM}_{n_p}\rangle$ with the threshold θ_α. If the included angle is larger than θ_α, it means that the BPDMs of the two pixels are dissimilar. In addition, if p is a root pixel, then its magnitude should be in the interval (d_{e_1}, d_{e_2}), and then insert the root pixel p into the set $\mathcal{R}$. Otherwise, the parent of p is updated to n_p. Because the root pixels in the same region are close to each other near the region's symmetry axis, parent $\mathcal{P}(r)$ should be updated to the last root pixel within the bottom half of 3×3 window centered at r. The final parent image $\mathcal{P}$ which represent initial segmentation is obtained via above operation.

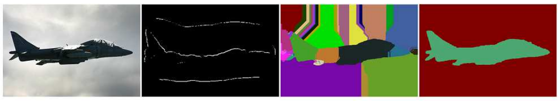
Algorithm 1 Generate optimized initial segmentation from the learned BPDMs (Sec.4)

Input: Learned BPDMs($\hat{DM}s$),Threshold($\theta_\alpha, d_{e_1}, d_{e_2}$)
Output: Parent image($\mathcal{P}$) and root pixel($\mathcal{R}$)

```
function Get_Initial_Seg($\hat{DM}s, \theta_\alpha, d_{e_1}, d_{e_2}$)
$\mathcal{P} \leftarrow p, \mathcal{R} \leftarrow \varnothing$
for each $p \in \Omega$ do
    if $cos^{-1}\langle \hat{DM}_p, \hat{DM}_{n_p}\rangle > \theta_\alpha$ and $d_{e_1} < \|\hat{DM}_p\| < d_{e_2}$ then
        $\mathcal{R} \leftarrow p$
    else
        $\mathcal{P}(p) \leftarrow n_p$
    end if
end for
for each $r \in \mathcal{R}$ do
    for each $q \in N_3^b(r)$ do
        if $q \in \mathcal{R}$ then
            $\mathcal{P}(p) \leftarrow q$
            $\mathcal{R}.pop(r)$
        end if
    end for
end for
return $\mathcal{P}, \mathcal{R}$
```

The pixels in the image can be combined into a forest of trees which are disjoint regions. As shown in Fig.4(b-c), the tree forest composed of these trees is the initial parent image, each tree has its corresponding root pixels.

Final segmentation. Similar to [21], for each pixel $r \in \mathcal{R}$, $\mathcal{A}_r$ is represent the area of the initial segment. Given the threshold α_s and α_t, the initial segments are divided into large, small and tiny regions and construct a region adjacency graphs $G(\mathcal{R}, E)$ based on the initial segmentation.



(a) Image (b) Root Pixels($\mathcal{R}$) (c) Initial($\mathcal{P}$) (d) Final

Fig. 4 The process of image segmentation. (a) input image, (b) root pixels, (c) initial segmentation, (d) final segmentation from initial segmentation by regions merged

The direction similarity $S(e)$ on each edge $e = (r_1, r_2) \in E$ which links two regions R_1 and R_2 is computed. $S(e)$ is defined as following:

$$S(e) = \pi - \frac{\sum_{i=1}^{|B(e)|} cos^{-1}\langle \hat{DM}_{\mathcal{P}s}(p_i), \hat{DM}_{\mathcal{P}s}(q_i)\rangle}{|B(e)|} \quad (9)$$

where $B(e) = \{(p_i, q_i)\}, p_i \in R_1, q_i \in R_2$ is defined as the pairs of boundary points between regions R_1, R_2, $|B(e)|$ is the numbers of the pairs. $\mathcal{P}_s(p)$ denotes the $s-th$ step starting from pixel.

If $S(e)$ is larger than given threshold h_θ which value is assigned according to the areas of adjacent regions, the two regions will be merged together. Through the above merge operation, small crumb regions in the initial segmentation can be cleaned up to the final segment result, as shown in Fig.4(d).

5 Experiments

5.1 Datasets

The performance of the presented algorithm is evaluated on three datasets of PASCAL Context [32], BSDS500 [33] and Cityscapes [34]. Pascal Context is a pixel level semantic annotation of the whole image and we re-labeled some obvious objects which are segmented as background in the dataset. 7072 images are used for training and 3031 images are used for testing.

BSDS500 includes 200 training sets, 100 verification sets and 200 test sets. Each image has about 5-10 ground-truth segmentations, and we select the finest ground-truth segmentation to train and test and expand the training set by rotating and flipping.

Cityscapes is a dataset of high-resolution urban image scenes, includes 2975 training images and 500 test images, in which every image has coarse label and fine label. In our experiment, fine label is used for supervised learning.

5.2 Training and Hyper-Parameters

In BPDM network, FCN adopts pretrained VGG16 on ImageNet to extract basic feature maps. During training model, the learning rate of the network is the same as [21] and optimizer uses ADAM [35]. The model is trained for 10000 epochs on each dataset respectively.

During initial segmentation, the hyper-parameters d_{e_1}, d_{e_2} are set to 2 and 23 respectively, and other hyper-parameters are same to [21].

All algorithms are trained and tested on 2xIntel Xeon Gold 6226R 16-core CPU (2.9GHz), 256GB RAM, and 4x NVIDIA Tesla V100S-PCIE-32GB GPU. The training of BPDM network is realized with Pytorch environment, and the final merging and segmentation is realized by using CUDA and C ++.

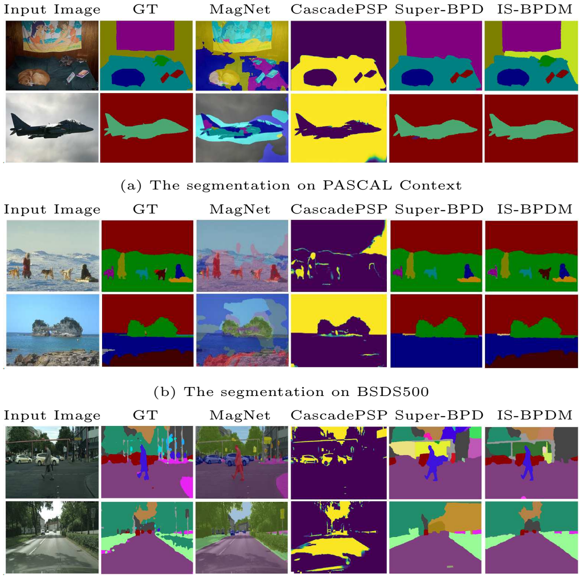
5.3 Qualitative and Quantitative Evaluations

To evaluate the performances of our method, mean Intersection over Union ($mIoU$) [16], F-measure for boundaries (F_b) [33] and computing expense are considered. $mIoU$ is used to assess the correlation between ground-truth and prediction, the higher the value, the better the performance of segmentation. Similarly, the higher F_b means the edge segmentation effect is better. For computational consumption, the values of time consuming are provided, and second is the unit.

The proposed IS-BPMD is tested and compared with some state-of-art segmentation methods such as CascadePSP [14], MagNet [15] and super-BPD [21], and colored the segmentation results of IS-BPDM and super-BPD by referring to ground truth. Some qualitative comparison results on the three datasets are shown in Fig.5. On Pascal Context and BSDS500 datasets, Our IS-BPDM can achieve better segmentation than MagNet, CascadePSP and super-BPD. It can segment overlapping objects (wall and painting) and small objects (books and bed) on PASCAL Context, and can clearly segment the dog and big stone on BSDS500. On Cityscapes dataset, although IS-BPDM does not obtain ideal result on both sides of the road segmentation than MagNet, it can segment overlapping people and cars, and also can segment small objects more finely than three other methods.

Table 1 In-dataset evaluation results. Ranking the top two indicators are bold.

Datasets	Methods	$mIoU$(%)	F_b	Time(s)
PASCAL Context	SILC[5]	48.45	0.419	**0.027**
	Mean Shift[4]	55.34	0.416	1.896
	Watershed[24]	-	0.667	0.057
	FCN-8s[16]	63.20	0.525	0.600
	MagNet[15]	64.70	0.688	0.580
	CascadePSP[14]	**70.23**	**0.740**	0.833
	Super-BPD[21]	69.15	0.731	**0.011**
	IS-BPDM(our)	**71.21**	**0.775**	0.039
BSDS500	SILC[5]	56.84	0.529	**0.023**
	Mean Shift[4]	61.34	0.608	2.543
	Watershed[24]	-	0.641	0.047
	FCN-8s[16]	66.75	0.647	0.632
	MagNet[15]	**70.40**	**0.703**	0.520
	CascadePSP[14]	70.20	0.698	0.712
	Super-BPD[21]	69.34	0.695	**0.010**
	IS-BPDM(our)	**76.40**	**0.724**	0.035
Cityscapes	SILC[5]	44.04	0.358	**0.026**
	Mean Shift[4]	49.12	0.488	14.477
	Watershed[24]	-	0.435	0.919
	FCN-8s[16]	55.41	0.518	0.700
	MagNet[15]	**67.57**	**0.688**	0.570
	CascadePSP[14]	65.34	0.644	1.795
	Super-BPD[21]	66.10	0.652	**0.023**
	IS-BPDM(our)	**67.17**	**0.668**	0.130



(a) The segmentation on PASCAL Context

(b) The segmentation on BSDS500

(c) The segmentation on Cityscapes

Fig. 5 Quantitative comparison between the proposed IS-BPDM method and other advanced methods on (a) PASCAL Context, (b) BSDS500, (c) Cityscapes dataset

Moreover, Table.1 illustrates the comparisons of our IS-BPDM algorithm and some widely used image segmentation methods on three datasets. On PASCAL Context and BSDS500 datasets, our IS-BPDM can learn the accurate BPDMs and make full use of the priori properties, so it can achieve the highest $mIoU$ and F_b, and has a good trade-off between accuracy and efficiency on segmenting images. On Cityscapes dataset in which images are high- resolution and their contents are complex, our IS-BPDM achieves the top two segmentation accuracy in $mIoU$ and F_b, it can effectively segment the house, people, car and so on, and also achieves a high efficiency.

It can be concluded that traditional segmentation methods rely on human intervention, and neural network approaches outperform traditional methods due to a large number training samples and strong fitting ability of the network. Our IS-BPDM approach uses neural network to learn accurate and robust BPDMs, and adopts traditional segmentation method with BPDMs to finish the finial segmentation according to the priori properties knowledge of BPDMs. The combinational model has a respectable segmentation and high efficiency.

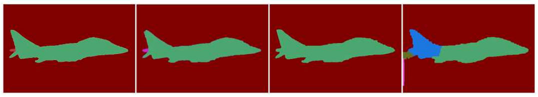
5.4 Ablation Studies

Module. We study the impact of adding watershed and attention mechanism modules to the network on Pascal Context. As stated in Table 2, when only watershed or attention mechanism is added, the segmentation performance is improved. Both watershed and attention mechanism modules achieve better results.

Table 2 The effects of watershed and attention mechanism modules on the performance in F_b and $mIoU$

Datasets	Watershed	Attention	F_b	$mIoU$
PASCAL Context			0.732	69.41
	✓		0.733	69.54
		✓	0.758	70.12
	✓	✓	0.775	71.21

Setting the n value of the adaptive weight. The importance of small regions during network training can be improved through adaptive weight $w(p)$ in Eq.(8). As shown in Fig.6, with the increase of n value, it is more sensitive to small regions, but this does not mean that the greater the n value, the better the segmentation. When n is less than 1, model pays more attention to the segmentation of large regions, so that the small regions cannot be segmented. When n is greater than 1, it pays too much attention to small regions, and results in over segmentation. When setting n to 1, we can achieve better segmentation performance.



(a)n=0 (b)n=0.5 (c)n=1 (d)n=1.5

Fig. 6 The influence of different n values on image segmentation

Direction characters. The effects of direction and magnitude characters on the initial segmentation based on BPDMs are depicted in Table.3. It can be seen that both direction and magnitude can achieve better segmentation results in comparison.

Table 3 The effects of direction difference on the performance in $mIoU$.

Datasets	Derection	Magnitude	$mIoU$
PASCAL Context	✓		70.13
	✓	✓	71.21

6 Conclusion and Future Work

For image segmentation, higher accuracy of edge and small regions and less time consuming are required. The proposed algorithm considers the pixels nearby boundary pixels in different regions should have opposite directions and shorter magnitude, and nearby root pixels in the same region have opposite directions and longer magnitude. So BPDM network which embed watershed and attention mechanism module is constructed and an adaptive loss function is used to train the BPDM network, which can effectively improve the accuracy and robustness of BPDMs on small areas and weak edges. Then the initial segmented regions are accomplished according to the pixel direction similarity and magnitude of BPDMs, and finally merge them into the final segmentation based on RAG. Experiments performed on PASCAL Context, BSDS500 and Cityscapes datasets show that the proposed IS-BPDM achieves a reasonable and accuracy performance for small object segmentation.

Though the proposed IS-BPDM is validated and outperforms a pleasant segmentation accuracy and efficiency, it still does not realize semantic segmentation. In the future work, we would like to consider the end-to-end semantic segmentation guided by BPDMs.

Acknowledgments. This work was supported by Foundation of Hubei Key Laboratory of metallurgical industry process system science (No.Y202008), National Natural Science Foundation of China(No.61671338).

References

1. Ghosh S, Das N, Das I, et al. Understanding deep learning techniques for image segmentation[J]. In CSUR, 2019, 52(4): 1-35.
2. Rimer S P, Mullapudi A, Troutman S C, et al. pystorms: A simulation sandbox for the development and evaluation of stormwater control algorithms[J]. arXiv preprint arXiv:2110.12289, 2021.
3. Zhao D, Liu L, Yu F, et al. Chaotic random spare ant colony optimization for multi-threshold image segmentation of 2D Kapur entropy[J]. Knowledge-Based Systems, 2021, 216: 106510.
4. Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. In PAMI, 2002 (5):603–619.
5. Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. SLIC superpixels compared to state-of-the-art superpixel methods. In PAMI,34(11):2274–2282, 2012.
6. Shi P , Zhong J , Rampun A , et al. A hierarchical pipeline for breast boundary segmentation and calcification detection in mammograms[J]. Computers in Biology and Medicine, 2018, 96:178.
7. Rother C, Kolmogorov V, Blake A. "GrabCut": interactive foreground extraction using iterated graph cuts[J]. Acm Trans. on Graphics,2004,23(3):309-314.
8. Caselles V, Kimmel R, Sapiro G. Geodesic active contours[C].In Proc. of CVPR. IEEE, 1995: 694-699.
9. Yu Y, Fang C, Liao Z. Piecewise flat embedding for image segmentation[C].In Proc. of ECCV. 2015: 1368-1376.

10. Kevis-Kokitsi Maninis, Jordi Pont-Tuset, Pablo Arbel´aez, and Luc Van Gool. Convolutional oriented boundaries: From image segmentation to high-level tasks. IEEE Trans. Pattern Anal. Mach. Intell, 40(4):819–833, 2018.
11. Jordi Pont-Tuset, Pablo Arbel´aez, Jonathan T Barron, Ferran Marques, and Jitendra Malik. Multiscale combinatorial grouping for image segmentation and object proposal generation. In PAMI, 39(1):128–140, 2017. 1, 2, 6, 7.
12. Ping Hu, Fabian Caba, Oliver Wang, Zhe Lin, Stan Sclaroff, and Federico Perazzi. Temporally distributed networks for fast video semantic segmentation. In CVPR, pages 8818–8827, 2020.
13. Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-toend object detection with transformers. In ECCV, 2020.
14. Cheng H K, Chung J, Tai Y W, et al. CascadePSP: toward class-agnostic and very high-resolution segmentation via global and local refinement[C].In Proc. of CVPR. 2020: 8890-8899.
15. Huynh C, Tran A T, Luu K, et al. Progressive Semantic Segmentation[C].In Proc. of CVPR. 2021: 16755-16764.
16. Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In Proc. of CVPR, 2015. 1, 8.
17. Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. In PAMI, 40(4):834–848, 2018.
18. Mishra, Puneet, and Ittai Herrmann. GAN meets chemometrics: Segmenting spectral images with pixel2pixel image translation with conditional generative adversarial networks. Chemometrics and Intelligent Laboratory Systems 215 (2021): 104362.
19. Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks. In Proc. of CVPR. 2017: 1125-1134.
20. Zhang X, Chen Z, Wu Q M J, et al. Fast semantic segmentation for scene perception[J]. IEEE Trans. Industr. Inform., 2018, 15(2): 1183-1192.
21. Jianqiang Wan, Yang Liu, Donglai Wei , Xiang Bai, Yongchao Xu. Super-BPD: Super Boundary-to-Pixel Direction for Fast Image Segmentation. In CVPR, 9250-9259, 2020.
22. Edwin Carlinet and Thierry Géraud. A comparative review of component tree computation algorithms. IEEE Trans. Image Process. 23(9):3885–3895, 2014.
23. Philippe Salembier, Albert Oliveras, and Luis Garrido. Antiextensive connected operators for image and sequence processing. IEEE Trans. Image Process. 7(4):555–570, 1998.
24. Yuan Y, Zhu Z, Yu H, et al. Watershed-based superpixels with global and local boundary marching[J]. IEEE Trans. Image Process. 2020, 29: 7375-7388.
25. Steffen Wolf, Constantin Pape, Alberto Bailoni, Nasim Rahaman, Anna Kreshuk, Ullrich Kothe, and FredA Hamprecht. The mutex watershed: efficient, parameter-free image partitioning. In Proc. of ECCV, pages 546–562, 2018. 2,6, 7.
26. Min Bai, Raquel Urtasun. Deep Watershed Transform for Instance Segmentation. In CVPR,2016.
27. Qu, Zhong, and Li Zhang. Research on image segmentation based on the improved Otsu algorithm. In IHMSC 2010. Vol. 2. IEEE, 2010.
28. Wang W, Song H, Zhao S, et al. Learning unsupervised video object segmentation through visual attention[C]. In Proc. of CVPR. 2019: 3064-3074.
29. Corbetta M, Shulman G L. Control of goal-directed and stimulus-driven attention in the brain[J]. Nature reviews neuroscience, 2002, 3(3): 201-215.
30. Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]. In Proc. of CVPR. 2018: 7132-7141.
31. Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]. In Proc. of ECCV. 2018: 3-19.
32. Roozbeh Mottaghi, Xianjie Chen, Xiaobai Liu, Nam-Gyu Cho, Seong-Whan Lee, Sanja Fidler, Raquel Urtasun, and Alan Yuille. The role of context for object detection and semantic segmentation in the wild. In Proc. of CVPR, pages 891–898, 2014.
33. Pablo Arbeláez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. In PAMI, 33(5):898–916, 2011.
34. Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In Proc. of CVPR, 2016.
35. Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Proc. of ICLR, volume 5, 2015. 5