Active Contour Building Segmentation Model based on Convolution Neural Network

Mengjia Liu\textsuperscript{1,2}, Peng Liu\textsuperscript{1}, Bingze Song\textsuperscript{1,2}, Yuwei Zhang\textsuperscript{1,2}, Luo Zhang\textsuperscript{1,2}

1. Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China
2. School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China
liupeng@radi.ac.cn

Abstract. In high-resolution remote sensing images, artificial features on the surface account for a large proportion. In artificial features, buildings, as special artificial features, buildings have different shapes. They are easily affected by light, so it takes a long time to extract using traditional image segmentation methods. It can't effectively design feature engineering to depict the high-dimensional features of the target building. We propose an active contour model based on a convolution neural network, which integrates the prior knowledge and constraints of active contour model, such as continuity of boundary, smooth edge, and geometric characteristics of buildings, into the learning process of convolution neural network to realize the close unity of ACM and CNN. According to our work, a fundamental end-to-end trainable image segmentation framework which is composed of convolution neural network (CNN) and ACM with learnable parameters is implemented, the problem of semantic segmentation of buildings in aerial images was dealt with, the model was evaluated on the publicly available dataset called Vaihingen, and some parameters were explained. In building semantics, the active contour model based on a convolution neural network has good performance.

Keywords: convolution neural network (CNN); active contour model (ACM); semantic segmentation

1. Introduction

In remote sensing, the development of image acquisition technology has always been ahead of the development of image processing technology. As remote sensing technology grows, obtaining remote sensing images is becoming more and more diversified. All kinds of military, civilian and commercial remote sensing satellites provide high-resolution, hyperspectral, and all-weather remote sensing data services for global users. In addition, the application field of remote sensing is also expanding, for instance, applied in the areas of land and resources exploration, agriculture production, forest protection, environment monitoring. At present, the amount of remote sensing image data is surging, the variety of remote sensing images is more abundant, and the remote sensing image acquisition technology is developing rapidly. However, the corresponding remote sensing image processing and key information extraction technology are relatively backward, slowing down the pace of remote sensing in people's daily lives [1].

Remote sensing image segmentation technology plays a vital role in remote sensing image digital processing, and it is the first step of many remote sensing image processing and analysis technologies. Effective segmentation guarantees deep feature extraction, such as remote sensing image analysis, recognition, information extraction, etc. The target detection of remote sensing images is easily affected
by their multiplicative speckle noise. The distribution of ground objects in some areas is complex, the contrast between target and background is low, and the target boundary is blurred in remote sensing images. The existence of these problems increases the difficulty of remote sensing image segmentation.

According to the research of Kim [12] et al., the CV model is used as a loss function, the parameter f is replaced by ground truth, tested in PASCAL VOC 2012 and Cityscapes, and FCN and DeepLab are used as a baseline, which ultimately increases the mIOU by 3% to 4%. Although the implementation of the model is simple, the reproduction effect is not good. The research of Chen [13] et al. has improved using the Convexified CV model, adding the regularization term related to length and achieving the effect of a 1% increase in Hausdorff distance. ACM is only designed as a loss function in the above research rather than part of an end-to-end model.

After 2010, artificial intelligence is becoming more and more popular. The neural network model is applied to computer vision and shines brilliantly in the field of computer vision. The new generation of image segmentation models generated by deep neural networks usually performs better than traditional image segmentation algorithms in undertaking and processing effects.

According to our work, an active co-contour model based on a convolution neural network is proposed, which combines the learning ability of a convolution neural network with people's prior knowledge of the target to be segmented (description of target shape, empirical statistics of brightness and color). Taking the aerial remote sensing image of urban architecture as an example, a semantic segmentation model for extracting dense urban buildings is constructed.

2. Related Work

Kass[2] et al. proposed an active contour model called Snake in 1987, based on global energy minimization. The model is guided by external constraints and pulls the initial contour to the target boundary under external rules. The internal energy term maintains the shape of the active contour.

Snake does not divide the areas used in the image, and it only divides specific regions or parts of particular areas. Snake relies on other mechanisms, such as interaction with users, higher-level image understanding processes, and information from image data adjacent to each other in time or space. This interaction must specify an approximate shape and initial profile of the starting position for the Snake somewhere near the target profile. Then, a priori and image-based information push the Snake to an appropriate energy minimization function. The minimized energy is a weighted combination of internal and external forces[3].

The active contour model depends on the deformable function controlled by the energy minimization function. The energy of the active contour relies not only on the image itself but also on the parameters of the deformable function and different prior knowledge. The energy of the active contour depends not only on the image itself but also on the parameters of the deformable function and additional prior knowledge:

The first and most important constraint is that because the goal of the active contour model is to perform segmentation based on object and shape detection, the active contour is required to be affected by force to make the active contour approach to the edge of the target object.

The second constraint depends on the profile model we want. We want to get a deformable model, which means that we want to have a continuous curve whose curvature will match the curvature of the target profile. Therefore, this means that there should be a theoretical model that we can use to describe the behavior of the active contour itself.

The energy of the active contour model is the sum of the internal energy term and the external energy term of the image[5]. Therefore, the energy function formula of the active contour model is obtained:

\[
E_{snake} = E_{int} + E_{ext} = \delta L \alpha \| \gamma'(s) \|^2 + \beta \| \gamma''(s) \|^2 ds - \delta f \| \nabla I \|^2 (\gamma(s)) ds \tag{1}
\]

As mentioned earlier, this active contour method depends on the minimization of the energy function. Now that the energy function is defined, the following work builds a way to minimize the energy function. The goal at this stage is to find the best parameters that reduce the defined energy
function. These parameters are position vectors that define the location of the active contour line, thus minimizing the energy function. The purpose of minimizing the energy function is to find the value of $\gamma(s) = (\gamma_x, \gamma_y)(s)$ when $E_{snake}$ is minimal.

Therefore, the optimization formula for this problem is:

$$S_{optimal} = \arg\min_{\gamma \in \mathcal{F}} E(\gamma(s))$$  \hspace{1cm} (2)

Infinite dimensions, one of the most used methods, is the gradient descent method. This method comes from function analysis because if the function's gradient is equal to zero, the function has reached the local extremum. The gradient descent method is a first-order optimization technique based on first-order Taylor series decomposition. Other higher-order decomposition methods, such as Newton's method, can also be used.

The active contour can be expressed as a discrete contour region $y = (up\ pence\ v)$, which contains $L$ control nodes, $y = (up\ pence\ v) \in \mathbb{R}^2$, where $s \in (1, L)$ each $s$ represents a control node of the discretized profile. Change the shape of the discrete contour region $y$ to minimize the following energy functions:

$$E(y, x) = \sum_{s=1}^{L} \left[ D(x, (y_s)) + \alpha(x, (y_s)) \frac{\partial y}{\partial s}^2 + \beta(x, (y_s)) \left[ \frac{\partial y}{\partial s}^2 \right]^2 + \sum_{u, v \in \Omega(y)} k(x, (u, v)) \right]$$  \hspace{1cm} (3)

Where $D(x) \in \mathbb{R}^{U \times V}$ is called data correlation, depending on the input image of $U \times V$ size, then $x \in \mathbb{R}^{U \times V}$, where $d$ is the number of image channels). $\alpha(x), \beta(x) \in \mathbb{R}^{U \times V}$ according to the introduction of the snake model in the second chapter, these two parameters make the discrete contour region tend to converge. Smooth, respectively. $k(x)$ is a balloon force term. The introduction of the balloon force term [6] can effectively prevent the discrete contour region from collapsing into a point where $\Omega(y)$ refers to the area surrounding the contour line. $D(x, (u_s, v_s))$ represents the value of $D(x)$ at the control point yard $y_s = (u_s, v_s)$. In the design of this model, $D, \beta, \text{and k are } U \times V$ vectors, and $\alpha$ is the scalar in our experiment.

2.1. Data Item

Data dependencies define the area where the active outline should be located. $D(x)$ is usually predefined function on an image, usually related to the image gradient. A suitable $D(x)$ should correspond to a relatively low value at the edge of the target object and a higher value in other areas. In the active contour model, we generally use the direction in which the gradient decreases the fastest, so we will:

$$-D(x) = -\left[ \frac{\partial D(x)}{\partial u}, \frac{\partial D(x)}{\partial v} \right]$$  \hspace{1cm} (4)

As a data correlation, the contour tends to the boundary of the target object.

2.2. Internal energy term

In the traditional active contour model, the values of $\alpha$ and $\beta$ are usually scalar, which means that there is the same feedback intensity in all image pixels, so we need to weigh the value of $\beta$ in the traditional active contour method. Changing the dimension of $\beta$ to assign different $\beta$ punishments to each pixel to avoid this tradeoff allows the precise edge information to be extracted effectively.

The function of the internal energy term is:

$$E_{int} = \alpha(x, (y_s))|y'|^2 + \beta(x, (y_s))|y''|^2$$  \hspace{1cm} (5)

Furthermore, we get the discretized form of the internal energy term:

$$E_{int} = \sum_{s=0}^{L} \alpha(y_s) \left| \frac{y_{s+1} - y_s}{\Delta s} \right|^2 + \beta(y_s) \left| \frac{y_{s+1} - 2y_s + y_{s-1}}{\Delta s^2} \right|^2$$  \hspace{1cm} (6)
thus the derivative of the internal energy term to the contour control point is calculated:

\[
\frac{\partial E_{\text{int}}}{\partial y_s} = \frac{2}{\Delta s} \left[ -\alpha_{s-1} \alpha_{s-1} + \alpha_s - \alpha_s \right] \cdot [y_{s-1}, y_s, y_{s+1}]^T + \frac{2}{\Delta s^2} \left[ \beta_{s-1}, -2\beta_s - 2\beta_{s-1}, \beta_{s-1} + 4\beta_s + \beta_{s+1}, -2\beta_{s+1} - 2\beta_s, \beta_{s+1} \right] \cdot [y_{s-2}, y_{s-1}, y_s, y_{s+1}, y_{s+2}]^T 
\]

transforming it into the form of an elegance-bit matrix:

\[
\frac{\partial E_{\text{int}}}{\partial y} = (A + B)y
\]

(8)

Where A (α) is a tridiagonal matrix, B (β) is a pentagonal matrix.

2.3. Balloon force term

The balloon force term refers to the expansion of the active contour by adding a constant external force in the average direction of each contour control node. Like the coefficient β, we increase its flexibility by taking a different value at each image location.

In the LD Cohen et al. [6], the balloon term is only considered the force increased after calculating the minima of other energy terms. In the model proposed in this paper, the balloon force term is regarded as part of the energy function related to the loss function we define.

Represent all balloon forces as a vector:

\[
n_s = [y_{s+1} - y_{s-1}] + 90^\circ = [v_{n+1} - v_{n-1}, u_{n+1} - u_{n-1}] 
\]

(9)

further expressed as:

\[
n = [Cv, u^TC] 
\]

Where C is a tridiagonal matrix, the main diagonal is 0, the upper diagonal is 1, and the lower diagonal is -1. Therefore, we can express the energy function of the balloon force term as:

\[
E_b = u^TCv = \int_{u, v \in \Omega(Y)} dudv 
\]

(10)

to flexibly control the balloon force, the coefficient \(k \in R^{M \times N}\) similar to \(\beta\) is added and discretized based on the original function:

\[
E_k = \sum_{u, v \in \Omega(Y)} k(u, v) 
\]

(11)

to obtain the derivative of the balloon force energy \(E_k\) to the vertical balloon force \(u_s\), we first obtain the relationship between \(\Delta E \) \(E_k\) and \(\Delta u_s\):

\[
\Delta E_k = \frac{\Delta u_s}{v_{s-1} - v_s} \int_{h=0}^{v_{s-1} - v_s} h_k(h)dh + \frac{\Delta u_s}{v_{s+1} - v_s} \int_{h=0}^{v_{s+1} - v_s} h_k(h)dh 
\]

(12)

therefore, the derivative of the balloon force term energy function concerning the vertical balloon force is:

\[
\frac{\partial E_k}{\partial u_s} = \frac{1}{v_{s-1} - v_s} \int_{h=0}^{v_{s-1} - v_s} h_k(h)dh + \frac{1}{v_{s+1} - v_s} \int_{h=0}^{v_{s+1} - v_s} h_k(h)dh 
\]

(13)

When we exchange \(u\) and \(v\) positions, we get the balloon force energy function derivative concerning the horizontal balloon force.

Finally, we obtain the iterative formula of the edge contour control points in the active contour model:

\[
y^{t+1} = y^t - \frac{\partial E_{\text{ext}}}{dy^t} - (A + B)y^t+1 
\]

(14)

further simplification to get:
\[ y^{t+1} = (I + A + B)^{-1}\left(y^t - \frac{dE_{ext}}{dy^t}\right) \] (15)

3. Network Architecture

3.1. Loss function

For each pair of input images and their corresponding ground truth images, it is expressed as \((y^i, x^i) \in y \times x, i = 1 \ldots N\), our goal is to minimize the deviation \(\Delta(y, \hat{y})\) between the predicted contour image and ground truth.

From the energy function in section 2, we can infer the expression of \(y\) times:

\[ \hat{y}^i = \arg\min_{y \in y} E(y, x, w) \] (16)

to minimize the \(\Delta(y, \hat{y})\), we turn the problem into:

\[ \hat{w} = \arg\min_{w} \sum_i \Delta\left(y^i, \arg\min_{y \in y} E(y, x, w)\right) \] (17)

because of the discontinuity of \(\Delta(y, \hat{y})\), we can replace it with a continuous convex function, such as the hinge loss function [7]. By adding L2 regularization and summing all the training samples, the maximum margin can be obtained. The formula is as follows:

\[ L(y, x, w) = \frac{1}{2} \|w\|^2 + C \sum_i \left(\max[0, \Delta(y, y^i) - E(y, x^i; w) + E(y^i, x^i; w)]\right) \] (18)

Since \(L(y, x, w)\) is a non-differentiable convex function, we need to calculate the sub-gradient [8], which requires us to determine that the current \(\omega\) is worthy of the maximum penalty constraint. In this case, the expressions of \(\hat{y}^i\) becomes:

\[ \hat{y}^i = \arg\max_{y \in y} \left[\Delta(y, y^i) - E(y, x^i; w)\right] \] (19)

the subgradient of \(w\) is expressed as:

\[ \frac{\partial L(y, x, w)}{\partial w} = w + C \sum_i \left(\frac{\partial E(y^i, x^i; w)}{\partial w} - \frac{\partial E(\hat{y}^i, x^i; w)}{\partial w}\right) \] (20)

therefore, we get the subgradients of \(D, a, \beta, k,\)::

\[ \frac{\partial \xi(y^i, x^i; w)}{\partial a(w(x^i))} = \left((u, v) \in y^i\right) - \left((u, v) \in \hat{y}^i\right) \] \quad (21)

\[ \frac{\partial \xi(y^i, x^i; w)}{\partial a(u(x^i))} = \left|\frac{\partial y(u, v)}{\partial s}\right|^2 \left((u, v) \in y^i\right) - \left|\frac{\partial y(u, v)}{\partial s}\right|^2 \left((u, v) \in \hat{y}^i\right) \] \quad (22)

\[ \frac{\partial \xi(y^i, x^i; w)}{\partial a(x^i)} = \left|\frac{\partial y(u, v)}{\partial s^2}\right|^2 \left((u, v) \in y^i\right) - \left|\frac{\partial y(u, v)}{\partial s^2}\right|^2 \left((u, v) \in \hat{y}^i\right) \] \quad (23)

\[ \frac{\partial \xi(y^i, x^i; w)}{\partial k(x^i)} = \left((u, v) \in \Omega(y^i)\right) - \left((u, v) \in \Omega(\hat{y}^i)\right) \] \quad (24)

Where [] is Iverson's parenthesis, which means that if the condition in the square bracket is met, it is 1, and if it is not satisfied, it is 0.
3.2. Network model

As shown in Figure 1, the CNN-based active model consists of two parts: the CNN module[4] and the ACM module. The CNN module learns four parameters of the energy function of the active contour model through the training set: the coefficient $\alpha$ related to the discreteness of the control points, the coefficient $D$ of the data correlation term, the coefficient $\beta$ associated with the smoothness of the discrete contour, and the coefficient $k$ related to the balloon force. The ACM part iterates continuously according to the initial contour to minimize the total energy and get the predicted contour. The structured loss function connects the CNN part with the ACM part and updates four main parameters by backpropagation. As shown in figure X. During each iteration, CNN is followed by ACM, which predicts a profile to calculate the structured loss, which is then sent back to CNN for backpropagation to update the parameters for the next iteration.

In the CNN module we designed, the input size of CNN is fixed, and the first six layers of the network are FCN networks, in which the convolution kernel size of the first layer is 7x7, the convolution core size of the second layer is 5x5. The convolution core size of the remaining four layers is 3x3. Each convolution layer is followed by ReLu, Batch normalization, and 2x2 pooling layer, and the output of each layer is sampled (or deconvolution) to the output size and spliced together. The number of convolution cores in the first six layers of the network is 32, 64, 128, 128, 256, 256, respectively. Figure 5.2 shows the network structure of the FCN part of the CNN module. Note that the pooling process and upsampling process are not shown in the diagram.

The FCN output is a three-dimensional vector, flattened into a one-dimensional vector, and input to a double-layer MLP with 256 and 64 hidden units to predict four output maps: $D(x), \alpha(x), \beta(x), k(x)$.

The ADAM optimizer with a learning rate of $10^{-4}$ is adopted in the CNN-based ACM model. We use geometric transformation for data enhancement. The number of iterations of the ACM module is set to 50.

4. Data and Experiment

In the experiment, we used the Vaihingen buildings dataset, which consists of 168 buildings extracted from the "2Dsemanticlabelingcontest" training sets[9][10][11]. The image has three bands corresponding to near-infrared, red and green wavelengths, and the resolution is 9cm. The Vaihingen data set uses high-resolution orthogonal photos and corresponding dense image matching technology to generate the digital terrain model (DSM). Vaihingen is a relatively small village with many separate buildings and small multi-story buildings. Each dataset has been manually classified into the six most common land cover categories. Vaihingen is a multi-category semantic segmentation data set, but we only need building categories in this experiment, so we need to process the original ground truth reference image.
As shown in Figure 2, the left is the original ground-truth reference image, and the right is the processed single building category image.

![Figure 2. multi-labels to single label](image1)

For remote sensing images, in the imaging process, the sensor will produce different images of different positions and shapes when the same object is photographed at different angles. Therefore, the image after geometric transformation can make the neural network model better learn the rotation invariance characteristics of ground objects to better adapt to varying forms of images. Therefore, we need to perform data enhancement operations on the dataset. Figure 3 shows the four states of an image: original image, horizontal flip, vertical flip, and diagonal flip.

The Bing huts data set and NWPU VHR-10-car data set are selected for training to test the model's performance in low-resolution images. The Bing huts data set consists of 605 individual huts, and aerial images with 30cm resolution are selected from rural Tanzania on Bing maps. Among them, 335 shots of 80 × 80 pixels are used to train the model, and the remaining 270 images are used for testing. Low spatial resolution, low contrast between buildings and surrounding soil, and high label noise make Bing huts a very challenging data set. NWPU VHR-10-car data set is an aerial remote sensing image set extracted from the NWPU VHR-10 data set, which only contains automobile class. Northwestern Polytechnical University released the NWPU VHR-10 data set in 2014. The color images of the data set were obtained from Google Earth with a spatial resolution range of 0.5m -2m. The extracted NWPU VHR-10-car has 339 images with a size of 128x128 pixel, of which 170 are used for training models and the other 169 are used for testing.

5. Result and Discussion

Table 1. Evaluation index of the model in each data set

|                           | Vaihingen | NWPU VHR-10-car | Bing huts |
|---------------------------|-----------|-----------------|-----------|
| Basic CNN                 | 0.77      | 0.55            | 0.56      |
| ACM based on CNN          | **0.83**  | **0.63**        | **0.65**  |
| ACM based on CNN (β, k as scalar) | 0.64  | -              | -         |
| ACM based on CNN (α as a vector) | 0.63  | -              | -         |
| ACM based on CNN (no k)   | **0.83**  | -              | -         |

*the symbol - means this value below 0.5*
Table 1 shows the average intersection and merges ratio ((mIoU)) of different models in different data sets. The ACM model based on CNN significantly improves the IOU benchmark. At the same time, the above other models confirm the necessity of \( \kappa \) and \( \beta \) non-scalar, while \( \alpha \) can be treated as a scalar without affecting the performance. It also emphasizes the importance of the balloon force term for contour convergence. The CNN-based ACM model is also improved by more than 5% on low-resolution data sets.

In Figure 4, the ground truth is a solid green line, the active outline is a yellow dashed line, and the initial outline is a blue dashed line. The example of the segmentation result of the Vaihingen data set shows that the parameters learned by the convolution neural network do produce smooth and straight edges while allowing sharp corners to exist. By looking at the energy terms predicted in Figure 5, we observe the following situations: The energy of the data correlation term \( D(X) \) is low on the boundary of the building, and the balloon force term \( \kappa(X) \) only produces a positive value in the interior of the building, especially near the corner of the building. In the curvature-related \( \beta(X) \), the weight near the corner of the building is low or even has no curvature.

In NWPUVHR-10-car and Bing huts datasets with low resolution and few pixels, the accuracy of the ACM model based on convolution neural network is 5% higher than that of the benchmark model. The above situation shows the potential of embedding advanced geometric features in the deep learning framework. The ACM model based on convolution neural network uses CNN to determine the parameters of the energy function of the active contour model (ACM) so that its output approaches the closed contour of the edge of the target object. By incorporating ACM reasoning into CNN training and using the creation of ACM and Ground Truth images to assess the structural loss, and then using backpropagation to update the CNN parameters for end-to-end training.

The ACM model based on a convolution neural network uses many energy item sets to match different data sets because the neural network automatically learns the balance between them. The main drawback of this model is that initialization is given through some external method, so it is not included in the learning process.
The ACM model based on a convolution neural network uses many energy item sets to match different data sets because the neural network automatically learns the balance between them. The main drawback of this model is that initialization is given through some external method, so it is not included in the learning process.

6. Conclusion
With the popularity of deep learning frameworks such as TensorFlow and PyTorch, neural network models have become more convenient and fast. The traditional image processing algorithms gradually fade out of the mainstream field of vision under the impact of neural networks represented by CNN. The conventional image processing algorithm is based on a solid theory. It can be explained by a perfect thesis, while the neural network model learns the characteristics of the image from the data set, which is abstract and difficult to explain. The interpretability of the model is theory-centered, and the model's defects are known and further improved through the complete explanation of the model so that human beings can trust the model. The lack of an interpretable network model is like a black box, which can only be found and patched through various black-box tests.

By combining the traditional image processing algorithm with the neural network, the universal standard image processing algorithm can be optimized in a specific field, and the significance of the neural network parameters can be inferred from the perfect theoretical basis of the traditional image processing algorithm.

References

[1] Haralick R M, Shapiro L G. Image segmentation techniques[J]. Computer vision, graphics, and image processing, 1985, 29(1): 100-132.
[2] Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models[J]. International journal of computer vision, 1988, 1(4): 321-331.
[3] Choi S, Kim C. Automatic initialization active contour model for the segmentation of the chest wall on chest CT[J]. Healthcare informatics research, 2010, 16(1): 36.
[4] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436-444.
[5] Kichenassamy S, Kumar A, Olver P, et al. gradient flows and geometric active contour models[C]//Proceedings of IEEE International Conference on Computer Vision. IEEE, 1995: 810-815.
[6] Cohen L D. On active contour models and balloons[J]. CVGIP: Image understanding, 1991, 53(2): 211-218.
[7] Vishwanathan S V M, Murty M N. SSVM: a simple SVM algorithm[C]//Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290). IEEE, 2002, 3: 2393-2398.
[8] Boyd S, Xiao L, Mutapcic A. Subgradient methods[J]. Lecture notes of EE392o, Stanford University, Autumn Quarter, 2003, 2004: 2004-2005.
[9] Gong Cheng, Junwei Han, Peicheng Zhou, Lei Guo. Multi-class geospatial object detection and geographic image classification based on collection of part detectors. ISPRS Journal of Photogrammetry and Remote Sensing, 98: 119-132, 2014.
[10] Gong Cheng, Junwei Han. A survey on object detection in optical remote sensing images. ISPRS Journal of Photogrammetry and Remote Sensing, 117: 11-28, 2016.
[11] Gong Cheng, Peicheng Zhou, Junwei Han. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 54(12): 7405-7415, 2016.
[12] Kim Y, Kim S, Kim T, et al. CNN-Based Semantic Segmentation Using Level Set Loss[C]//2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2019: 1752-1760.
[13] Chen X, Williams B M, Vallabhaneni S R, et al. Learning Active Contour Models for Medical
Image Segmentation[C]/Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 11632-11640.

**Acknowledgment**
This work was supported by the National Natural Science Foundation of China under Grant 61731022, Grant 41971397.