Segmentation Quality Evaluation Algorithm Based On Classification

Ting Feng, Bo Peng, Zaid Al-Huda and Hao Jia

September 26, 2021

Segmentation Quality Evaluation Algorithm Based On Classification

1st T. Feng
School of Computing and Artificial Intelligence
Southwest Jiaotong University
Chengdu, China
tfeng@my.swjtu.edu.cn

2nd B. Peng
School of Computing and Artificial Intelligence
Southwest Jiaotong University
Chengdu, China
bpeng@swjtu.edu.cn

3rd Z. Al-Huda
School of Computing and Artificial Intelligence
Southwest Jiaotong University
Chengdu, China
zaid.alhuda@my.swjtu.edu.cn

4th H. Jia
School of Computing and Artificial Intelligence
Southwest Jiaotong University
Chengdu, China
jiahao@my.swjtu.edu.cn

***Abstract*—Image segmentation plays a vital role in computer vision. Image segmentation quality evaluation is an essential task in image segmentation and an important step to quantify the performance of the segmentation algorithm. Most of the existing evaluation methods need to use ground truth to evaluate the segmentation quality. However, the annotation of ground truth varies from person to person and takes a long time. In this paper, we proposed a new segmentation quality evaluation network. In the evaluation of segmentation quality, only the segmentation results to be evaluated and the original image are required without specially annotated ground truth. At the same time, we also propose a new space to squeeze module (STS) for segmentation quality evaluation. STS module autonomously learns the edge features of the segmentation object and increases the weight of edge features. Experiments on the dataset constructed in this paper show that the performance of the proposed network is better than other network structures such as ECA-Net, CBAM, SE-Net, and the evaluation accuracy is higher than the existing supervised and unsupervised segmentation quality evaluation methods.**

***Index Terms*—segmentation evaluation, network model, unsupervised, classification**

I. Introduction

Image segmentation [1] is a fundamental problem in the field of computer vision, which is the premise of image analysis and understanding. In recent years, researchers have proposed many image segmentation algorithms, hoping that different scenes, different environments, different types of image segmentation can be close to human interpretation of the image. Although image segmentation algorithms have been widely studied with the emergence and development of various segmentation algorithms, there is little research on the quantitative evaluation of image segmentation quality. The primary purpose of segmentation quality evaluation is to measure the quality of image segmentation results, improve [2] the performance of the segmentation algorithm, and promote the rapid development of image processing. It is the post-processing step of image segmentation.

Image segmentation quality evaluation methods [3] are mainly divided into the subjective evaluation and objective evaluation. The main difference between the two methods is whether the observer participates in the evaluation. Human beings play a key role in image segmentation and assessment. The subjective evaluation method is considered the most reliable evaluation method. Still, the subjective method is very time-consuming and requires observers of different backgrounds and ages to participate in the evaluation. Even if the segmentation is visually close, different observers may give further evaluations. The Objective evaluation method aims to accurately and automatically predict the quality of segmentation results, which is more and more popular.

Objective evaluation methods are divided into supervised evaluation methods based on ground truth and unsupervised evaluation methods without ground truth. In the supervised evaluation method, we need to label the ground truth manually. The evaluation method is to calculate the overlap between the segmentation result and the ground truth region. Different people may label the ground truth differently, which has particular subjectivity and takes a long time to mark manually. In [4], local features and corners of the segmentation result graph are used to judge the similarity of ground truth. When ground truth labeling is wrong, misjudgment results are likely to occur. In addition, the supervised evaluation method ignores the texture [5], semantic information [6] and color [7], and only considers the edge information [8] and regional features [9] of the segmentation result. Ground truth can not describe the complex semantic information of the image.

Although many unsupervised [10] evaluation methods have been proposed, the accuracy and robustness of these methods can not meet the actual needs. Haralick et al. [11] proposed some good segmentation criteria, and Zhang et al. [10] summarized them as semantic case, the difference between regions and consistency within regions. The current unsupervised evaluation methods mainly follow the differences between regions and the consistency within areas to design segmentation quality indicators. However, the contrast between regions and within regions is only one aspect of measuring the quality of segmentation, and the evaluation effect is poor in images with complex texture information. Using CNN to extract more semantic and contextual information [12] from images can improve the accuracy and reliability of segmentation quality evaluation. Still, there are few unsupervised evaluation methods based on CNN. Huang et al. [13] regarded segmentation quality evaluation as a regression problem, proposed three kinds of segmentation quality evaluation networks based on the semantic features of segmented images, and obtained the final quality score by using IOU and prediction score. The IOU ignores the semantic information of the image and is calculated by the intersection and union ratio of the segmentation result mask and the ground truth object region. The IOU obtained by the poor segmentation result may be more prominent, and the IOU obtained by the excellent segmentation result may be smaller.

In this paper, we regard the quality evaluation of image segmentation as a binary classification [14] problem and use 0-1 as the label of sample images to train images with different segmentation quality. Among them, 1 represents a good segmentation result, and 0 illustrates a poor segmentation result. According to the classification results, the classification probability is used as the quality score to evaluate the segmentation quality. The main contributions of this paper are as follows:

- In this paper, a new segmentation evaluation network is proposed. The network does not need ground truth but only needs to learn the features of the original image and the image to be evaluated. According to the learned edge features of the object, the segmentation results of other object classes can be assessed unsupervised.
- In this paper, a new STS module for segmentation quality evaluation is proposed. By encoding and decoding the plane space of image features, the module actively learns the weight of features in the space. It gives more weight to essential edge features to achieve the complete extraction of object edge features.
- Because of the lack of open object segmentation evaluation data set, this paper uses all the pictures in the VOC2012 data set as the original picture to construct a large-scale object segmentation evaluation data set. The data set includes 20 object classes, and the negative samples are various, which can be applied to most segmentation quality evaluation cases.

II. Related Works

At present, many image segmentation quality evaluation methods have been proposed. In this part, we mainly introduce supervised evaluation methods and unsupervised evaluation methods.

The supervised evaluation method evaluates the performance of the image segmentation algorithm by comparing the similarity between image segmentation results and ground truth. The similarity between segmentation results and ground truth determines the quality of segmentation results. However, ground truth is manually segmented by humans, and human segmentation mainly depends on visual perception ability. Different perception abilities and focus, different ground truth, and supervised evaluation methods have inevitable subjectivity. At present, the commonly used supervised evaluation methods include: SC [12], Dice [15], cross union ratio (IOU), probability rand index (PRI) [16] and information variation (VI) [17] , etc. These evaluation methods usually use the color, histogram, and other design image features to evaluate the segmentation results by measuring the similarity between the segmentation results and ground truth.

The unsupervised evaluation method does not need ground truth, which is the only way to evaluate the quality of image segmentation online. Unsupervised evaluation methods usually meet three criteria:

- The semantic characteristics of the measurement object.
- The difference between the measurement regions.
- The consistency within the measurement regions.

Early unsupervised evaluation methods usually only focus on gray images, and commonly used indicators include $F^{'}$, Q, Zeb, Ecw, F, E, etc., summarized by Zhang et al. [10]. F calculates each segmented region's average color square error and penalizes over-segmentation by giving weight proportional to the square root of the total number of segmented regions. F calculates the variance of the color in the area. The smaller the minor F is, the smaller the change of the color value in the region is, the better the segmentation result. Because F tends to over-segmentation, it produces more small regions than expected. $F^{'}$ is an improvement of F. $F^{'}$ improves the deviation of F by punishing many small regions of the same size. Because $F^{'}$ tends to be the unsegmented image. Q improves $F^{'}$ and uses fewer regions to represent all the objects in the image, which reduces the tendency of over-segmentation and non-segmentation. Zeb is an unsupervised evaluation method based on intra-region comparison and inter-region comparison of each pixel. The visual error between regions is used to evaluate the over-segmentation. It is helpful to assess the under segmentation by the visual error in the part. E is an evaluation function based on information theory and minimum description length (MDL). It uses region entropy to measure the uniformity of the region, that is, to measure the intensity of each pixel in the area. When the region entropy decreases, the layout entropy (the entropy of which region the pixel belongs to) is used to punish the over-segmentation of the image. Although there is no precise measurement standard for the difference between regions, it is implied in the combination of region entropy and layout entropy. The over-segmentation and under-segmentation can be balanced by mutual inhibition of the two kinds of entropy.

When the image has complex texture information or objects in complex scenes, the above methods have certain limitations. Due to the subjectivity of the supervised evaluation method, the ground truth artificial segmentation is wrong, and the

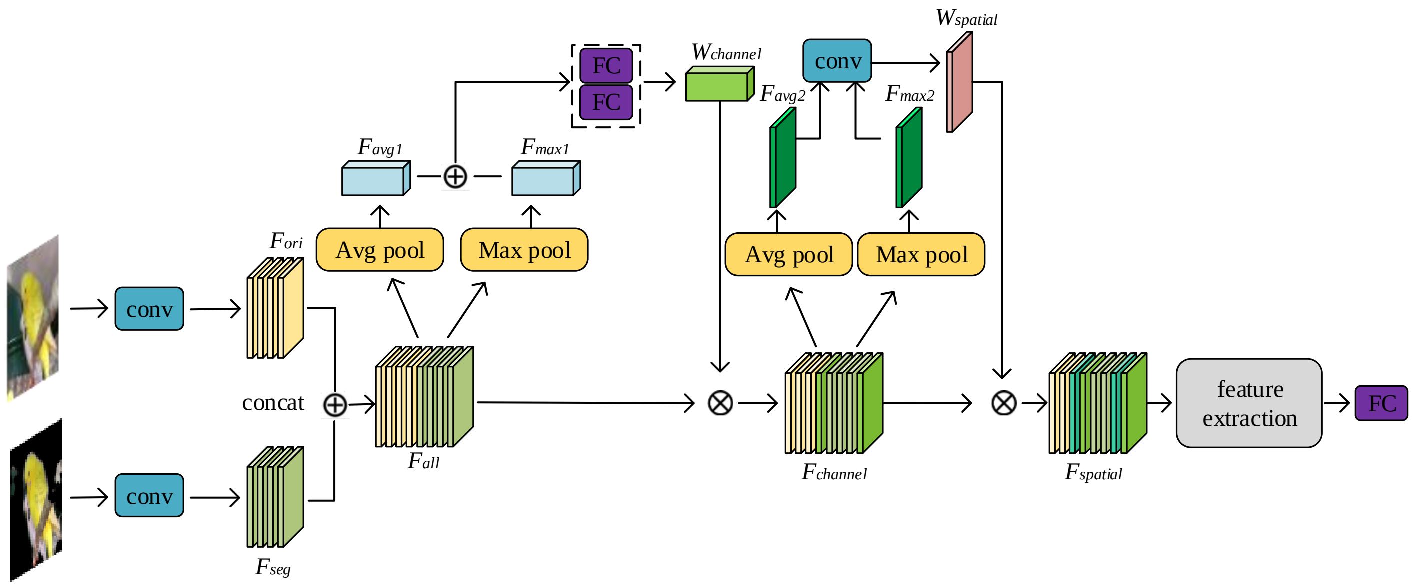

Fig. 1. Network structure diagram of segmentation quality evaluation framework.

incorrect evaluation results are obtained. However, the early unsupervised evaluation method uses a single intra-region feature or inter-region feature to evaluate the segmentation results. When the image has complex semantic information, the evaluation effect has a specific error. For the first time, this method uses channel attention, spatial attention, and the importance of image plane spatial edge features to carry out autonomous learning and increase the weight of essential edge features for image segmentation evaluation to achieve the total extraction of important edge features. Binary classification is a question of "whether or not" and the evaluation of segmentation quality corresponds to binary classification. The idea of binary classification can be used to evaluate the quality of segmentation. A good evaluation should make the evaluation result of good segmentation result (positive sample) close to 1 and that of poor segmentation result (negative sample) close to 0. According to the classification results, this paper uses the classification probability as the quality score to evaluate the segmentation results, which is feasible in practical application.

III. Method

This part first introduces the segmentation quality evaluation network proposed by us. Secondly, in order to extract the important edge features of the image segmentation results, the end-to-end segmentation results are evaluated accurately, and the STS module is presented. Finally, the construction process of object segmentation evaluation data set is introduced.

A. Network Structure

To construct a convolutional neural network (CNN) suitable for most object segmentation quality evaluation, we regard the image segmentation quality evaluation as a binary classification problem. Use ResNet18 [18] as the backbone network. To obtain more global and local information of the image, double branch 3×3 convolution is used to extract the features of the original image and the segmentation results, and get $F_{\text{ori}} \in R^{H\times W\times C}$ and $F_{\text{seg}} \in R^{H\times W\times \boldsymbol{C}}$. H represents the height of feature F, W represents the width of feature F, and C represents the channel number of F. The feature fusion of F_{ori} and F_{seg} is carried out according to (1). The channel of the fused feature map is 16, and the feature map $F_{\text{all}} \in R^{H\times W\times C}$ is obtained. For getting the salient features on the channel from F_{all}, F_{avg1} and F_{max1} are obtained by performing average pool and max pool operations on F_{all} respectively. F_{avg1} and F_{max1} are added to get the feature weight W_{channel} on the channel through two FC operations. F_{channel} is obtained by multiplying W_{channel} and F_{all}. Next, in order to get the spatial salient features from F_{channel}, average pool and max pool operations are performed on F_{channel} to get F_{avg2} and F_{max2}, convolution operations are performed on F_{avg2} and F_{max2} to get the spatial feature weight W_{spatial}, and F_{spatial} is obtained by multiplying W_{spatial} and F_{channel}. In the evaluation of segmentation quality, let the network pay more attention to the salient features of the image. At the same time, the STS module is connected to ResNet18's block to extract the features of F_{spatial}. The loss function used in the network model is cross entropy loss. In the training phase, weight is assigned to each class.

$$F_{\text{all}} = F_{\text{ori}} \odot F_{\text{seg}} \tag{1}$$

Where $\odot$ means F_{ori} and F_{seg} are connected by channel.

To fully extract the critical features of the original image and the segmentation result image in an end-to-end way, the

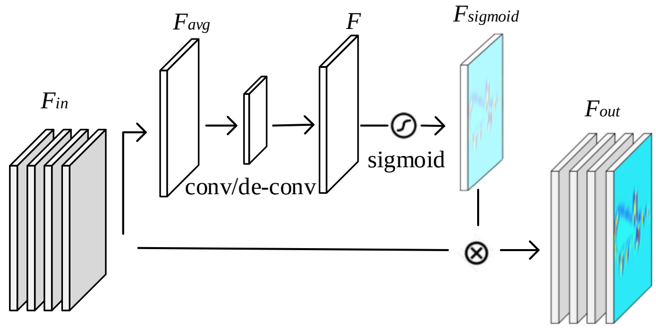

Fig. 2. The structure diagram of STS module.

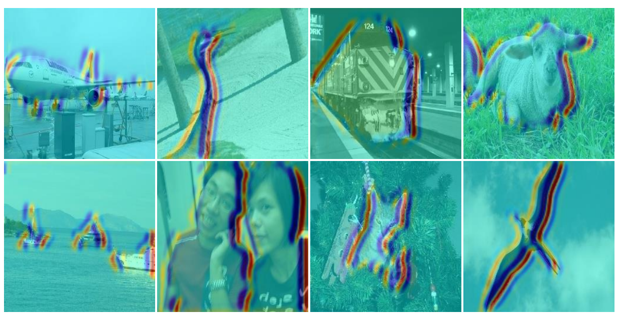

Fig. 3. The segmentation evaluation network is added to the visualization of the first layer of STS.

unsupervised evaluation of the segmentation result is realized. This paper proposes an STS module, which uses convolution and deconvolution to learn the weight of all features in the image by module. It gives a large weight to the important edge features of the object in the image. In order to avoid the important edge features being treated as common features equally, it does not have an obvious distinction. Firstly, avg-pool is performed on input feature F_{in} to get F_{avg}, Conv and ConvTranspose are performed on F_{avg} to get feature F, and then Sigmoid activation function is used to activate F according to (2):

$$F_{\text{sigmoid}} = \frac{1}{1+e^{-F(x)}} \tag{2}$$

The characteristic graph $F_{\text{sigmoid}} \in R^{H\times W}$ is obtained. The larger value in F corresponds to the larger weight in $F_{\text{sigmoid}} \in R^{H\times W}$. In this paper, through the self-learning of image feature weight, we give the important edge features in the image to be evaluated greater weight, highlight the important edge features of the object, and improve the accuracy of segmentation quality evaluation. For each plane space of input feature F_{in}, the final feature vector is calculated according to (3):

$$F_{\text{out}} = F_{\text{in}} \otimes F_{\text{sigmoid}} \tag{3}$$

Where, every plane space of F_{in} is multiplied by F_{sigmoid}, and the specific flow is shown in the structure diagram of STS module in Fig. 2.

This paper visualizes the network after the segmentation quality evaluation network is added to STS, and the results of the first layer of convolution layer are shown in Fig. 3. The network uses the STS module to learn the weight of essential edge features, which can extract the critical edge features of the object and evaluate the quality of segmentation accurately.

The segmentation evaluation method proposed in this paper only needs to extract the original image and the segmentation results to be evaluated and then classify the object eigenvector. According to the classification results, the quality of segmentation can be quantitatively judged by the probability obtained. The evaluation of arbitrary segmentation results can be carried out without the need for ground truth. The evaluation score of the objects from the same class is calculated according to (4):

$$\text{Score}(Xi) = \begin{cases} \max(\text{Softmax}(Xi)), & \text{if } argmax\ (Xi) = 1 \\ \min(\text{Softmax}(Xi)), & \text{if } argmax\ (Xi) = 0 \end{cases} \tag{4}$$

Where Xi is the feature vector of the object segmentation result, which $argmax(Xi) = 1$ means that the classification result is a positive sample, and $argmax(Xi) = 0$ implies that the classification result is a negative sample.

B. Object Segmentation Evaluation Dataset

Different types of objects have additional semantic information, and segmentation quality evaluation is very sensitive to the change of semantic information. Therefore, an object segmentation evaluation data set should contain enough object categories to ensure the comprehensiveness of segmentation evaluation. Due to the lack of public data set for segmentation quality evaluation, we select all the images in the PASCAL VOC2012 data set as our original image. The PASCAL VOC2012 data set contains 20 object class images. The original image includes a single object and multiple objects and has a complex or straightforward context, reflecting the comprehensiveness of the image segmentation algorithm in daily scenes.

The specific process of generating segmentation results is as follows: four mainstream image object segmentation algorithms (FCN [19], U2Net [20], UNet [21], DeepLab V3 [22]) are used to generate the segmentation results of each class. Different segmentation results can be obtained using different segmentation algorithms, which can reflect the objective diversity of segmentation results. For FCN, the 15th epoch model is selected to generate the segmentation result of each class; U2Net determines the 50th epoch model to create the segmentation result of each class; UNet and DeepLab V3 use the model saved in the 4th and 30th epoch respectively to generate the segmentation results of each type; Four different segmentation results are generated for each image as candidate images of the data set.

We determine the positive and negative sample labels in the data set according to the following methods:

- Positive samples contain only one class of objects.
- Negative samples can contain multiple classes, negative samples can be segmentation results of other images, and negative samples can be ground truth of different images.
- In this paper, according to the method of meta-evaluation [23], the segmentation results meet (5) to determine the positive and negative samples, where S1, S2, and S3 are the GT and two different segmentation results of the same graph respectively. The similarity between positive samples and GT is significant, while that between negative samples and GT is small. Because the positive samples are relatively single, only the object is segmented, and a convolutional neural network can accurately fit the distribution of positive samples. To accurately evaluate a variety of negative samples, we expand the negative samples in the data set, including the segmentation results of other images and the ground truth of other images in addition to the negative samples determined by (5).

Finally, we get a train set: 5029 positive samples and 10088 negative samples, of which 5029 negative samples corresponding to positive samples, 5059 segmentation results, and ground truth of other images are obtained; Test set: 571 positive samples and 1112 negative samples, including 571 negative samples corresponding to positive samples, 541 segmentation results, and ground truth of other pictures. As shown in Fig. 4, some datasets are shown.

$$|M(S1) - M(S2)| < |M(S1) - M(S3)| \tag{5}$$

IV. Experiment

A. Experimental Configuration

We chose ResNet18 as the backbone network. The original image and segmentation result are resized to 256×256, and each RGB channel of the image is normalized. The optimizer uses Adam, the learning rate is set to 0.001, the weight_decay is set to 0.0005, betas is set to 0.9 $\sim$ 0.99, and the batch size is set to 16. All models are trained in 50 epochs. The experimental hardware platform is: Intel(R) Core(TM) i7-9700F CPU @ 3.00GHz(GeForce RTX 2080Ti) 12GB; the software environment is: Ubuntu 16.04, Pytorch1.2.0.

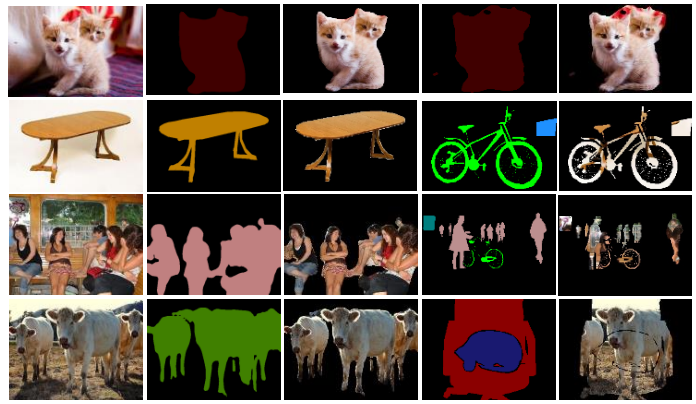

Fig. 4. The sample samples of the object segmentation evaluation data set, a is the original image, b is positive sample mask, c is a positive sample, d is negative sample mask, e is a negative sample.

B. Experimental Comparison

This part is used to verify the segmentation quality evaluation network on the object segmentation quality evaluation data set constructed in III-B.

C. Comparison With Other Networks

To evaluate the classification performance of our proposed segmentation evaluation method on the data set constructed in III-B; Three kinds of widely used ResNet18, ResNet34, and ResNet50 are used as backbone networks. The proposed method is compared with the three most advanced networks, including ECA-Net, CBAM, and SE-Net. It can be seen from Table I that the classification accuracy of our proposed method is 2.76% higher than that of the original ResNet18, TPR is the highest on ResNet18, 93.52%, and TNR is the highest on ECA-Net, 95.23%. On ResNet34, the TPR and ACC of our method are the highes, 91.59%, and the TNR of ResNet34 is the highest, 93.26%. On ResNet50, the TPR of ECA-Net is the highest, 89.67%, and the TNR of ResNet50 is the highest, 96.67%. Although the TPR and TNR of our method are not all the highest in the three backbone networks, the difference between TPR and TNR is smaller than that of the other four networks. The evaluation of positive and negative samples is more stable. The reason why our method has better performance is that we use convolution and deconvolution for reference. The critical edge features in the segmentation result are given more weight by the weight of self-learning features to avoid the extraction of essential edge features with the same importance as common features. ResNet classification network extracts all elements with the same priority, which has certain limitations in image segmentation quality evaluation.

D. Compared With the Unsupervised Evaluation Index

The proposed method does not need ground truth and can evaluate different segmentation results. We compare the proposed method with the six evaluation indexes of unsupervised

TABLE I
CLASSIFICATION ACCURACY COMPARISON RESULTS.(%).

Method	Backbone	TPR	TNR	ACC
ResNet	ResNet18	**93.52**	83.81	88.67
ECA-Net		84.59	**95.23**	89.91
CBAM		82.66	93.71	88.19
SE-Net		76.88	87.86	82.37
Ours		89.14	93.71	**91.43**
ResNet	ResNet34	79.68	**93.26**	86.47
ECA-Net		83.89	92.36	88.13
CBAM		88.97	79.95	84.46
SE-Net		83.54	77.43	80.49
Ours		**91.59**	89.84	**90.72**
ResNet	ResNet50	74.78	**96.67**	85.73
ECA-Net		**89.67**	84.44	87.06
CBAM		87.57	83.72	85.65
SE-Net		86.54	89.25	87.90
Ours		86.69	95.05	**90.87**

TABLE II
ACCURACY COMPARISON RESULTS OF UNSUPERVISED EVALUATION METHODS.(%).

Method	testset
$F^{'}$	63.9
Q	85.9
Zeb	81.6
Ecw	59.2
F	62.3
E	75.6
Ours	**91.43**

evaluation method summarized by Zhang et al. [10], which are $F^{'}$, Q, Zeb, Ecw, F, E. These six unsupervised evaluation indexes measure the quality of segmentation by calculating the similarity within or between regions of the image and ignore the semantic information and texture information of the segmentation results. There are some limitations in the evaluation of segmentation quality. To accurately compare our proposed method with the six selected unsupervised evaluation methods, we carry out the accuracy comparison experiment on the test set. It can be seen from Table II that the highest accuracy of our proposed method is 91.43%, and only Q and Zeb are more than 80% of the unsupervised evaluation indexes. The main reason for the performance difference is that the traditional unsupervised evaluation method ignores the image's semantic information and texture information. When the image has complex texture information and semantic information, the early unsupervised evaluation measures the quality of the segmentation result using the intra-region consistency and inter-region difference of the image, and there are some errors. We propose a segmentation evaluation method, which uses channel attention and spatial attention, to pay attention to the critical information of images from both channel and space, which can improve the accuracy of the evaluation to a certain extent. Secondly, the STS module proposed in this paper uses convolution to extract the segmentation results and the features of the original image. It then uses deconvolution to increase the extracted feature scale. Then, by self-learning, the sigmoid activation function gives greater weight to the vital edge feature information to distinguish common features. Even if the image has complex texture information, the method can accurately evaluate the segmentation results.

E. Compared With the Supervised Evaluation Method

The mainstream segmentation quality evaluation method is still supervised evaluation [24] method. The supervised evaluation method usually ignores semantic information and texture information. When the segmentation result image has complex texture information, it can't be evaluated accurately. Moreover, the supervised evaluation method relies on the artificial segmentation of ground truth. When the artificial segmentation of ground truth is wrong, the supervised evaluation method will make an evaluation error when evaluating the segmentation quality. The proposed evaluation method does not need ground truth and can evaluate the segmentation results online. In order to verify whether our proposed evaluation method can accurately assess all kinds of segmentation results, we compare it with SC [12], Dice [15], cross union ratio (IOU), probability rand index (PRI) [16], and information variation (VI) [17] in the supervised evaluation method, and conduct accuracy comparison experiments on the test set constructed in III-B. The experimental results are shown in Table III: the accuracy of our proposed method is the highest, exceeding 90%, while the accuracy of IOU in the supervised evaluation index is the highest, 87.29%. When the image has complex texture information, the image to be evaluated is the segmentation result of other images or the ground truth of different images. The supervised evaluation method only evaluates the segmentation result by measuring the similarity between the segmentation result and the ground truth. Because the ground truth ignores the semantic information of the image, and there will be artificial segmentation differences. Therefore, the supervised segmentation evaluation index only uses the region overlap degree, which can not accurately evaluate the segmentation

TABLE III
ACCURACY COMPARISON RESULTS OF SUPERVISED EVALUATION METHODS.(%).

Method	testset
SC	83.24
Dice	75.51
IoU	87.29
PRI	76.58
VI	82.98
Ours	**91.25**

quality. Compared with the supervised evaluation method, the proposed method fully considers the importance of image edge features in segmentation results. Even if the image to be evaluated has complex texture information, and the image to be assessed is the segmentation result or ground truth of other images, our proposed evaluation method can make full use of the edge information of the image to evaluate the segmentation quality.

In addition, the proposed method only needs to train the same class of objects in advance and evaluate the quality of segmentation results of all kinds of objects. To verify the generalization of our method, we test the accuracy of segmentation quality evaluation on the DUTS-TR dataset, and the accuracy of assessment is 87.18%. Therefore, this method can be extended and applied to the evaluation task without ground truth, and the segmentation result evaluation accuracy is high.

F. Qualitative Contrast Experiment

Most image object segmentation algorithms use IOU to measure the quality of segmentation results. To more intuitively verify whether our proposed method can quantitatively evaluate the quality of segmentation results, we qualitatively compare it with IOU. The results of comparative experiments are shown in Fig. 5: from the 2nd row, we can see that the IOU value of negative samples is close to 0.7. That is to say, and negative samples are regarded as positive samples, and misjudgment occurs. Our method gives negative samples a lower score, which is different from positive samples. From the 3rd and 4th rows, we can see that the negative sample is obtained using the mask of other image segmentation results, which belongs to image segmentation error. However, the IOU value of the negative sample is about 0.5, which means that the segmentation of a certain area of the image is correct. Our method outputs a higher score for the positive sample, close to 1, and a lower score for the negative sample, which is close to 0. It can clearly distinguish the quality of segmentation and accord with the results of human eye evaluation. According to (6), IOU measures the quality of segmentation result by the overlapping degree of segmentation result and ground truth region. When the ground truth manual annotation is wrong, a good segmentation result may have a small overlapping degree with the ground truth region. The IOU obtained is smaller, while the IOU corresponding to a good segmentation

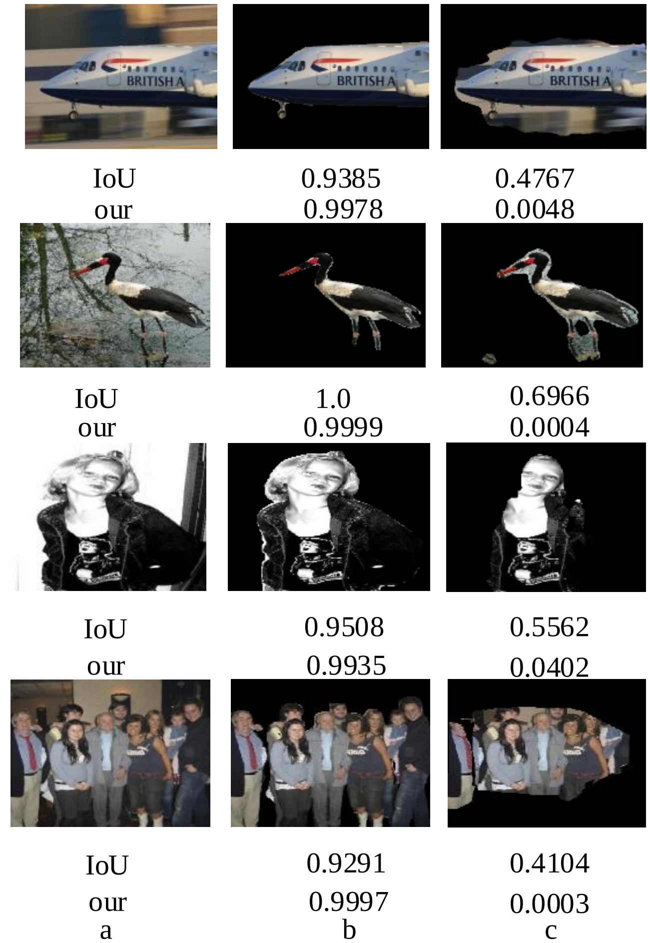

Fig. 5. a is the original image, b is the score of positive samples, c is the score of negative samples..

result is larger in practice. The poor segmentation results may overlap with the ground truth region. The IOU is larger, while the IOU corresponding to the poor segmentation results is smaller; when other images' segmentation results overlap the ground truth area, the evaluation error will also occur. In this case, using the size of the IOU value to measure the quality of the segmentation result will lead to the situation of misjudgment of the segmentation result. Instead of using region overlap, we use CNN embedded STS module to give important edge feature information to be evaluated a larger weight. Without using ground truth, the network can extract significant edge features of segmentation results. When the quality of segmentation results is good or bad, it can be accurately evaluated according to the critical edge features.

The calculation formula of IOU is as follows:

$$IoU = \frac{GT_i \cap Seg_i}{GT_i \cup Seg_i} \tag{6}$$

Where GT_i is the ground truth of the region, which Seg_i is the result of region segmentation.

CONCLUSION

In this paper, we propose a new segmentation quality evaluation network. To use more semantic information in the original image and the segmentation results to be evaluated, we use double branch convolution to extract the features of the original image and the segmentation results. The STS module proposed in this paper only needs to learn the features of the original image and the image to be evaluated. According to the learned object edge features, it can perform the unsupervised evaluation on the segmentation results of other object classes. Secondly, because of the lack of an open data set, we construct a data set of segmentation evaluation that contains a single object, multiple objects, and complex information according to the method of meta-evaluation. The data set has various negative samples, which can be applied to most segmentation quality evaluation cases. In addition, the performance of the proposed network is compared with ECA-Net, CBAM, and SE-Net. The experimental results show that the understanding of the proposed network is better than the other three networks. We compare the proposed unsupervised evaluation method with the existing supervised and unsupervised evaluation method. The experimental results show that our proposed method has high accuracy in evaluating the segmentation quality and conforms to the human eye evaluation results.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation of China (Nos. 61772435, 61961038), Sichuan highway science and technology project (No. 2019-01) and the Fundamental Research Funds for the Central Universities (No.2682021ZTPY069).

REFERENCES

[1] L. Fan, H. Kong, W. Wang and J. Yan, "Semantic Segmentation With Global Encoding and Dilated Decoder in Street Scenes," in IEEE Access, vol. 6, pp. 50333-50343, 2018, doi: 10.1109/ACCESS.2018.2868801.

[2] H. Li, F. Meng, B. Luo and S. Zhu, "Repairing Bad Co-Segmentation Using Its Quality Evaluation and Segment Propagation," in IEEE Transactions on Image Processing, vol. 23, no. 8, pp. 3545-3559, Aug. 2014, doi: 10.1109/TIP.2014.2330759.

[3] Wang Z , Wang E , Zhu Y . Image segmentation evaluation: a survey of methods[J]. Artificial Intelligence Review, 2020(1).

[4] M. Simfukwe, B. Peng and T. Li, "H2: Fusion of HOG and Harris features for image segmentation evaluation," 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), 2017, pp. 1-6, doi: 10.1109/ISKE.2017.8258832.

[5] Y. Chen, D. Dai, J. Pont-Tuset, and L. Gool. Scale-aware alignment of hierarchical image segmentation. In IEEE Conference on Computer Vision and Pattern Recognition, pages 364–372, 2016.

[6] Christensen H I , Phillips P J . Empirical Evaluation Methods in Computer Vision[M]. World Scientific, 2002.

[7] M. Meila. Comparing clusterings: an axiomatic view. In International Conference on Machine Learning, pages 577–584, 2005.

[8] R. Unnikrishnan, C. Pantofaru, and M. Hebert. A measure for objective evaluation of image segmentation algorithms. In Workshop on Empirical Evaluation Methods in Computer Vision, IEEE Conference on Computer Vision and Pattern Recognition, page 34, 2005.

[9] S. Wang. New benchmark for image segmentation evaluation. Journal of Electronic Imaging, 16(3):033011, 2008.

[10] Hui Zhang, Jason E. Fritts, Sally A. Goldman, Image segmentation evaluation: A survey of unsupervised methods, Computer Vision and Image Understanding, Volume 110, Issue 2, 2008, Pages 260-280, ISSN 1077-3142, https://doi.org/10.1016/j.cviu.2007.08.003.

[11] Robert M Haralick and Linda G Shapiro. Image segmentation techniques. Computer vision, graphics, and image processing, 29(1):100–132, 1985.

[12] P. Arbelaez, M. Maire, C. C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(5):898–916, 2011.

[13] Chao, H. , Q. Wu , and F. Meng . "QualityNet: Segmentation quality evaluation with deep convolutional networks." Visual Communications & Image Processing IEEE, 2016.

[14] Xue, Y., Xu, T., Zhang, H. et al. SegAN: Adversarial Network with Multi-scale L1 Loss for Medical Image Segmentation. Neuroinform 16, 383–392, 2018.

[15] L. Dice. Measures of the amount of ecologic association between species. Ecology, 26(3):297–302, 1945.

[16] R. Unnikrishnan, C. Pantofaru, and M. Hebert. A measure for objective evaluation of image segmentation algorithms. In Workshop on Empirical Evaluation Methods in Computer Vision, IEEE Conference on Computer Vision and Pattern Recognition, page 34, 2005.

[17] P. Viola and W. Wells. Alignment by maximization of mutual information. International Journal of Computer Vision, 24(2):137–C154, 1997.

[18] He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian. Deep Residual Learning for Image Recognition. CVPR, June, 2016.

[19] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In IEEE Conference on Computer Vision and Pattern Recognition, pages 3431–3440, 2015.

[20] Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar R. Zaiane, Martin Jagersand, U2-Net: Going deeper with nested U-structure for salient object detection,Pattern Recognition, Volume 106, 2020, 107404, ISSN 0031-3203.

[21] Ronneberger, O. , P. Fischer , and T. Brox . "U-Net: Convolutional Networks for Biomedical Image Segmentation." International Conference on Medical Image Computing and Computer-Assisted Intervention Springer International Publishing, 2015.

[22] L. Chen, G. Papandreou, F. Schroff, and H. Adam. Rethinking Atrous Convolution for Semantic Image Segmentation. Technical report, Google LA, Dec 2017.

[23] J. Pont-Tuset and F. Marques. Measures and meta-measures for the supervised evaluation of image segmentation. In IEEE Conference on Computer Vision and Pattern Recognition, pages 2131–2138, 2013.

[24] J. Pont-Tuset and F. Marques. Supervised evaluation of image segmentation and object proposal techniques. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(7):1465–1478, 2016.