



# Article A Semantic Segmentation Algorithm Using FCN with Combination of BSLIC

Wei Zhao<sup>1</sup>, Haodi Zhang<sup>1</sup>, Yujin Yan<sup>1</sup>, Yi Fu<sup>1</sup> and Hai Wang<sup>2,\*</sup>

- Key Laboratory of Electronic Equipment Structure Design, Ministry of Education, Xidian University, Xi'an 710071, China; weizhao@xidian.edu.cn (W.Z.); haodiz@stu.xidian.edu.cn (H.Z.); yanyujin.xidian@gmail.com (Y.Y.); yfu@stu.xidian.edu.cn (Y.F.)
- <sup>2</sup> School of Aerospace Science and Technology, Xidian University, Xi'an 710071, China
- \* Correspondence: wanghai@mail.xidian.edu.cn; Tel.: +86-029-8820-3115

Received: 12 February 2018; Accepted: 23 March 2018; Published: 26 March 2018



**Abstract:** An image semantic segmentation algorithm using fully convolutional network (FCN) integrated with the recently proposed simple linear iterative clustering (SLIC) that is based on boundary term (BSLIC) is developed. To improve the segmentation accuracy, the developed algorithm combines the FCN semantic segmentation results with the superpixel information acquired by BSLIC. During the combination process, the superpixel semantic annotation is newly introduced and realized by the four criteria. The four criteria are used to annotate a superpixel region, according to FCN semantic segmentation result. The developed algorithm can not only accurately identify the semantic information of the target in the image, but also achieve a high accuracy in the positioning of small edges. The effectiveness of our algorithm is evaluated on the dataset PASCAL VOC 2012. Experimental results show that the developed algorithm improved the target segmentation accuracy in comparison with the traditional FCN model. With the BSLIC superpixel information that is involved, the proposed algorithm can get 3.86%, 1.41%, and 1.28% improvement in pixel accuracy (PA) over FCN-32s, FCN-16s, and FCN-8s, respectively.

Keywords: semantic segmentation; FCN; BSLIC; superpixel

# 1. Introduction

Research on image semantic segmentation has been well-developed over decades [1–9]. Image semantic segmentation is a process in which an image is divided into several non-overlapping meaningful regions with semantic annotations being labeled. In an early study, the full-supervised semantic segmentation is developed by probabilistic graphical models (PGM), such as the generative model [1] and the discriminative model [10]. These models are based on the assumption of conditional independence, which might be too restrictive for many applications. To solve this problem, the condition random field model (CRF) [11] is proposed, which allows for us to get the correlation between variables. Moreover, the CRF model has the ability to exploit the local texture features and the contextual information. However, the CRF model can't acquire the overall shape feature of the image, which might cause misunderstanding during the analysis of a single target. Currently, one of the most popular deep learning techniques for semantic segmentation is the fully convolutional network (FCN) [4]. Unlike the traditional recognition [12–17] and segmentation [18,19] methods, FCN can be regarded as a CNN variant, which has the ability to extract the features of objects in the image. By replacing the fully connected layers with the convolutional layers, the classical CNN model, such as AlexNet [20], VGGNet [21], and GoogLeNet [22], is transformed into the FCN model. When compared with the CNN-based methods, the obtained model can achieve a significant improvement in segmentation accuracy. Despite the power and flexibility of the FCN model, it still has some problems

that hinder its application to certain situations: small objects are often ignored and misclassified as background, the detailed structures of an object are often lost or smoothed.

In 2003, Ren et al. [23] proposed a segmentation method, named as superpixel. A superpixel can be treated as the set of pixels that are similar in location, color, texture, etc. It reduces the number of entities to be labeled semantically and enable feature computation on bigger, more meaningful regions. There are several methods to generate superpixel, such as Turbopixels [24], SuperPB [25], SEEDS [26], and SLIC [27]. By comparison, SLIC is significantly more efficient and accurate than others. In the previous work, BSLIC is presented [28,29] for better superpixel generation that is based on SLIC. BSLIC can generate superpixels with a better trade-off between the regularity of shape and the fitting degree of the edge, so as to keep the detailed structures of an object.

In this paper, we propose a semantic segmentation algorithm by combining FCN with BSLIC. FCN is typically good at extracting the overall shape of an object. However, FCN ignores to focus on the detailed structures of an object, which still has room to improve. Meanwhile, BSLIC is used to efficiently align the SLIC superpixels to image edges and to obtain better segmentation results. In contrast to SLIC, BSLIC adopts a simpler edge detecting algorithm to obtain the complete boundary information. To combine FCN semantic segmentation result and BSLIC superpixel information, we newly introduce the superpixel semantic annotation, which is realized by four criteria. With the four criteria, the tsuperpixel region is annotated by FCN semantic segmentation result. By this means, the detailed structures of an object can be kept to some extent. The proposed algorithm can not only extract the high-level abstract features of the image, but also take full advantages of the detail of the image to improve the segmentation accuracy of the model.

The rest of the paper is organized as follows: In Section 2, the theory that is related to the proposed algorithm is introduced, including the fully convolution network and the BSLIC superpixel segmentation. The details of the algorithm in this paper are described in Section 3, and the experimental results are given in Section 4. In Section 5, an overall conclusion is made.

### 2. Related Work

#### 2.1. FCN

As a CNN variant, FCN is widely used for the pixel-level classification. The output of FCN is a category matrix with the semantic information of the image. The core ideas of the FCN model are summarized: (1) converting a fully connected layer to a convolution layer with the kernel size  $1 \times 1$ ; (2) using a deconvolution or upsampling operation to generate a category matrix with the same size as the input image. By this means, the popular FCN-8s, FCN-16s, and FCN-32s are defined in [4], and the generation procedure of them is described in Figure 1.



Figure 1. Generation of the fully convolutional network (FCN)-8s, FCN-16s and FCN-32s.

The 1st row of Figure 1 shows that the feature maps are derived from the input image by repeated convolution calculation and pooling operation. The convolution calculation is used to extract features which cannot change the resolution of the image or feature maps. Each pooling operation can reduce the resolution of the feature map by two times. Therefore, the resolution of the *n*th (n = 1, 2, ..., 5) feature map is  $1/2^n$  of the input image. Based on these feature maps, the semantic segmentation result, with the same resolution of the input image, can be obtained by using upsampling and multiscale fusion techniques. The upsampling is introduced to increase the resolution of feature map. Multiscale fusion is used to acquire finer details from the feature maps of different scale. Three different combinations of upsampling and muliscale fusion are utilized to implement FCN-8s, FCN-16s, and FCN-32s. As shown in Figure 1, the feature maps used to generate the semantic segmentation results are from pool3, pool4 and conv7, which are 1/8, 1/16, and 1/32 resolution of the input image, separately. The detailed description is as follows:

• FCN-32s

The conv7 feature map is upsampled 32 times directly to obtain the semantic segmentation result.

• FCN-16s

The conv7 feature map is upsampled 2 times firstly. Then the two-time-upsampled conv7 feature map and pool4 feature map are fused by using the mutiscale fusion technique. Finally, the fusion result is upsampled 16 times to obtain the semantic segmentation result.

• FCN-8s

The conv7 feature map is upsampled four times firstly. Then, the pool4 feature map is upsampled two times. Next the four-time-upsampled conv7 feature map, two-time-upsampled pool4 feature map, and pool3 feature map are fused by using the multiscale fusion technique. Finally, the fusion result is upsampled eight times to obtain the semantic segmentation result.

# 2.2. BSLIC

BSLIC comes from SLIC, which is a method of to generate superpixel. Essentially, BSLIC is a local K-means clustering algorithm. Assuming that the number of pixels in the given image is N, the number of superpixels to be generated is k, the coordinates of the pixels in the image are x and y and the components of CIELAB color space are l, a, and b, the SLIC algorithm is proposed as follows:

• Step 1:

The clustering center  $C_i = [l_i a_i b_i x_i y_i]^T$  is uniformly initialized in horizontal and vertical directions separately. The horizontal interval  $S_h$  and vertical interval  $S_v$  are as follows:

 $S_h = \sqrt{\frac{N}{\left(\frac{\sqrt{3}}{2}\right) \times k}},\tag{1}$ 

$$S_v = \sqrt{\frac{\sqrt{3}}{2} \times \frac{N}{k}},\tag{2}$$

# • Step 2:

In order to avoid selecting noise points in the image, all of the clustering centers are updated to points that have the smallest gradient value in their neighborhood.

# • Step 3:

In the  $2S_0 \times 2S_0$  neighborhood of each image pixel *i*, the clustering center with the smallest distance to the pixel *i* is searched to achieve the class tag of the pixel *i*.

### • Step 4:

The clustering center is updated to the mean vector of all the pixel eigenvectors in its category.

### • Step 5:

Repeat the third and fourth steps until the last two iterations of the cluster center error is no more than 5%.

BSLIC has an improvement in three aspects: initializing cluster centers in hexagon, rather than square distribution, additionally choosing some specific edge pixels as cluster centers, incorporating boundary term into the distance measurement during *k*-means clustering.

Figure 2 shows the segmentation results of BSLIC algorithm in different input parameters k and m, where k is the number of superpixels to be generated, and m is the weighting factor of color euclidean distance and the space euclidean distance. Through evaluating the segmentation results, we can draw the following conclusions: (1) larger m achieves better regularity for superpixels, and smaller m achieves better adherence to image boundaries; and, (2) the larger k aligns to more detailed boundaries.



**Figure 2.** The segmentation results of BSLIC algorithm under different input parameters. The image resolution is  $481 \times 321$ . From the first row to the third row, the parameter superpixels number *k* is 100, 500, and 1000. From the first column to the third column, the parameter weight factor *m* is 5, 15, and 30.

### 3. Proposed Algorithm

### 3.1. Overall Framwork

To make the FCN model achieve a more accurate and detailed description of the target edge, this paper integrates the superpixel edge of the image into the FCN model. The block diagram of our proposed algorithm that is based on FCN and BSLIC superpixel is shown in Figure 3. First, the FCN model based on VGG-16 network is trained on the dataset PASCAL VOC 2012 in order to develop the FCN semantic segmentation model. Then, the trained FCN semantic segmentation model is used to segment the image to obtain the image pixel-level classification map. At the same time, BSLIC superpixel segmentation is performed on the segmented image, generating a superpixel segmentation. BSLIC superpixel is semantically annotated by the pixel-level classification map, and the classification of the image at the pixel level is obtained, which is the revised semantic segmentation result. The pixel-level classification map combines the advantages of high-level semantic information and good edge information.



Figure 3. Flow chart of image semantic segmentation algorithm based on FCN and BSLIC.

### 3.2. Superpixel Semantic Annotation

Based on the framework mentioned above, superpixel semantic annotation (Algorithm 1) combines the FCN pixel-level classification map and the BSLIC superpixel segmentation map, and it finally gets the revised semantic segmentation result.

Algorit	hm 1. Superpixel s	emantic annotation			
1.	Acquire a FCN pixel-level classification map.				
2.	Acquire a BSLIC superpixel segmentation map.				
	The parameters are shown as follows:				
	the collection of superpixel $Sp = \{Sp_1, Sp_2, \cdots, Sp_k, \cdots, Sp_K\},\$				
	the number of semantic categories in superpixel $Sp_k$ is A, the proportion of pixels of the semantic category $t(0, 1, \dots, 20)$ ,				
	all pixels in the superpixel $Sp_k$ is $C_t$ .				
	Generate superpixel semantic annotation using the four criteria				
	<b>Loop</b> : For $k = 1:K$				
3	Criterion 1	If there is no image edge in superpixel $Sp_k$ and $A = 1$ then			
5.		Label the superpixel with FCN semantic result			
		End			
	Criterion 2	If there is no image edge in superpixel $Sp_k$ and $A > 1$ then			
		Use <i>t</i> of the largest $C_t$ to label the superpixel			
		End			
	Criterion 3	If there is image edge in superpixel $Sp_k$ and $A = 1$ then			
		Label the superpixel with FCN semantic result			
		End			
	Criterion 4	If there is image edge in superpixel $Sp_k$ and $A > 1$ then			
		If $C_t > 80\%$ in superpixel $Sp_k$ then			
		Use <i>t</i> of the largest $C_t$ to label the superpixel			
		Else			
		Maintain then FCN semantic segmentation result			
	End				
4.	Output the superpixel semantic annotation result.				

The core of the Superpixel semantic annotation is the four criteria, which can simplify the process of analyzing and calculating. According to the four criteria, the combination of FCN and BSLIC can be classified into four situations. Based on this simple and complete classification, the corresponding solutions are designed to combine FCN and BSLIC. All the four criteria are described as follows:

# • Criterion 1:

Situation: there is no image edge in the superpixel, and the FCN semantic of all pixels are the same; Solution: use the FCN pixel-level classification map to label all the pixels in this superpixel.

# • Criterion 2:

Situation: there is no image edge in the superpixel, but the FCN semantic of the pixel is different; Solution: use the category of the largest proportion  $C_t(t = 0, 1, 2, 3, ...)$  to label all of the pixels in this superpuixel.

# • Criterion 3:

Situation: there are image edges in the superpixel, but the FCN semantic of all the pixels are the same;

Solution: use the FCN pixel-level classification map to label the pixels in this superpixel.

# • Criterion 4:

Situation: there are edges in the superpixel, and the FCN semantic of the pixels are different. Solution 1: if there is any proportion  $C_t$  more than 80%, and then label all of the pixels in this superpixel as the category of the largest proportion  $C_t$ .

Solution 2: if there is no proportion  $C_t$  more than 80%, then maintain the FCN pixel-level classification result.

As shown in Figure 4, an example of superpixel semantic annotation is given to explain how to use the four criteria. Based on the superpixel segmentation map (Figure 4a) obtained by BSLIC, the detailed explanation is discussed as follows.



**Figure 4.** Example of superpixel semantic annotation. (**a**) The superpixel segmentation map; (**b**) region 1 in (**a**); (**c**) region 2 in (**a**); and, (**d**) region 3 in (**a**).

In Figure 4b, the situation of superpixel  $Sp_{219}$ , which is the local area of the aeroplane, meets criterion 1. The pixels of superpixel  $Sp_{219}$  (447 pixels) are labeled as aeroplane (category 1) in the FCN pixel-level classification map, so the pixels in this superpixel can be labeled as aeroplane directly.

In Figure 4c, the situation of superpixel  $Sp_{400}$ , which is the junction of wing and background, meets criterion 2. FCN model labels 75.5% of the pixels in the superpixel as background, and labels 24.5% of the pixels as aeroplane. Therefore, the classification of the pixels in this superpixel is background.

In Figure 4d, the situation of superpixel  $Sp_{178}$ , which is the area of fuselage, meets criterion 3. FCN model labels the pixels in this superpixel as aeroplane. Therefore, the pixels in this superpixel can be labeled as aeroplane directly.

In Figure 4b, the situation of superpixel  $Sp_{233}$ , which is the fuzzy edge of the image, meets criterion 4. 499 pixels (91.6%) in this superpixel are labeled by FCN as background. Consequently, the superpixel can be labeled as background directly.

#### 4. Experimental Results

Experiments are performed to evaluate the proposed algorithm in this section. The analysis of the experimental data will be given in detail. The experimental platform for this paper is: Intel 6700K@4.00GHz CPU, 32G memory, Samsung 840Pro SSD, Windows7 x64 operating system, Matlab 2014a development platform, MatConvNet depth learning toolbox beta23 version.

#### 4.1. Training and Testing of FCN Model

The used dataset in our work is PASCAL VOC 2012 with a total of 17,125 images. All of the targets are divided into 20 categories, excluding the background labeled as category 0. At the same time, in order to facilitate the expression of image semantics, PASCAL VOC 2012 assigns specific color labels and numbers to each category so that different categories of targets can be distinguished by color or numbers in image semantic segmentation. Figure 5 shows 20 categories of instance images and corresponding color labels. The dataset will be used to train and test the FCN network, and the detailed processed is shown as follows:

First, divide the PASCAL VOC 2012 dataset. The dataset involves 20 foreground object categories and one background category, as shown in Figure 5. The original dataset contains less than 5000 pixel-level labeled images. The extra annotations provided by [30] have 10,582 pixel-level labeled images. We take 9522 (90%) images as the training dataset, 530 (5%) images as the validation dataset, and the remaining 530 (5%) images as the test dataset.



Figure 5. PASCAL VOC 2012 dataset of 20 categories, numbers, and color labels.

Secondly, reconstruct the VGG-16 network for FCN. By converting the fully connected layers in the VGG-16 network into the convolution layers, a FCN-VGG16 model is acquired. After that, the FCN model is initialized by using the full 16-layer version, which is trained by OverFeat [21].

Thirdly, in order to have a better training result, we run the FCN for 50 epochs on training dataset. At the same time, we use pixel-level label to supervise the FCN-VGG16 model. The training model classification accuracy over the epoch of training is shown in Figure 6. The line that is marked by 'train' shows the relationship between the classification accuracy of training dataset with the epoch. The line marked by 'val' shows the relationship between the classification accuracy of the validation dataset with the epoch. Due to the randomness of the model parameter initialization, the classification accuracy of the validation dataset may be higher than that of the training dataset at initial epochs. Furthermore, the knowledge learning from the training dataset cannot suit the validation dataset well, so the curve 'val' becomes lower than the curve 'train' over epoch, leading to an intersection point around certain epochs.

Finally, validation dataset has been used to evaluate the training performance. The semantic segmentation results are obtained from FCN-8s, FCN-16s, and FCN-32s, respectively. Some of the experimental results are shown in Figure 7. The successful training model is used for semantic segmentation in further process.



Figure 6. FCN-VGG16 model classification accuracy over epochs of training.



**Figure 7.** Comparison of original images, semantic segmentation results and ground truth. (**a**) Original images; (**b**) FCN-32s semantic segmentation results; (**c**) FCN-16s semantic segmentation results; (**d**) FCN-8s semantic segmentation results; and, (**e**) the ground truth.

### 4.2. Qualitative Comparison

Due to the outstanding performance, FCN-8s is selected to be combined with BSLIC in our work. Our proposed algorithm is compared with three other methods: FCN-8s, FCN-16s, and FCN-32s. All of the test dataset from PASCAL VOC 2012 are processed by four algorithms. Owing to the better performance observed in Figure 7, FCN-8s is selected as the reference for comparison. Figure 8 compares the FCN semantic segmentation results with our semantic segmentation results. From Figure 8c,d, it is observed that the proposed algorithm can reach the same semantic recognition accuracy as FCN-8s. At the same time, the improved algorithm has a better accuracy when dealing with the small edges of the target, such as the wheel of aeroplane, the bird's feet, the sheep's ear, and the rear view mirror of a car in Figure 8d. In FCN-8s, small objects are often ignored and misclassified as background. In conclusion, the combination of FCN and BSLIC outperforms FCN-8s, FCN-16s and FCN-32s, especially the ability to keep the detailed structure of an object.



Figure 8. Cont.



**Figure 8.** Combination FCN and BSLIC to acquire more details of a target. (a) The original test images; (b) The BSLIC superpixel segmentation results; (c) The FCN-8S semantic segmentation results; (d) The results of the improved model proposed in this paper; and, (e) The ground truth.

Table 1 shows the time-consuming of four methods. The proposed method allows for FCN and BSLIC to run at the same time, so there is no time to waste. However, it requires a little extra time to combine FCN and BSLIC by using the four criteria. This is the reason why the four criteria are designed in a simple and efficient way.

Table 1. Time-consuming of four methods.

Method	Average Time (ms/image)		
FCN-32s	293.050		
FCN-16s	297.358		
FCN-8s	302.796		
Ours	310.218		

#### 4.3. Quantitative Comparison

In practical engineering, pixel accuracy (PA), intersection over union (IoU), and mean intersection over union (mIoU) are used to evaluate the performance of the semantic segmentation technique. PA is used to measure the accuracy of object contour segmentation. IoU is used to measure the accuracy of an object detector on a particular dataset. mIoU is the average of IoU and is defined to reflect the overall enhancement of semantic segmentation accuracy. The formulas to calculate metrics can be found in [31] and is used in this paper. It is assumed that the total of classes is (k + 1) and  $p_{ij}$  is the amount of pixels of class *i* inferred to class *j*.  $p_{ii}$  represents the number of true positives, while  $p_{ij}$  and  $p_{ij}$  are usually interpreted as false positives and false negatives, respectively.

• Pixel Accuracy (PA):

It computes a ratio between the amount of properly classified pixels and the total number of them. The formula can be expressed as:

$$PA = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{i=0}^{k} \sum_{j=0}^{k} p_{ij}},$$
(3)

### • Intersection Over Union (IoU):

The IoU is used to measure whether the target in the image is detected. The formula can be expressed as:

$$IoU = \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}},$$
(4)

# • Mean Intersection Over Union (mIoU):

This is the standard metric for segmentation purposes. It computes a ratio between the ground truth and our predicted segmentation. mIoU is computed by averaging IoU.

$$mIoU = \frac{1}{k+1} \times \sum_{k} IoU_k,$$
(5)

Based on the previous test dataset from PASCAL VOC 2012, PA, IoU, and mIoU are calculated by (3)–(5), respectively.

Table 2 gives the IoU scores between SegNet [32], UNIST\_GDN\_FCN [33], and our proposed method. Table 3 gives the IoU scores of FCN-32s, FCN-16s, FCN-8s, and our method.

Category	SegNet	UNIST_GDN_FCN	Ours
0	90.3 *	89.8	89.7
1	73.6	74.5	74.7 *
2	37.6 *	31.9	36.9
3	62.0	66.7	77.4 *
4	46.8	49.7	52.4 *
5	58.6	60.5	67.7 *
6	79.1 *	76.9	75.5
7	70.1	75.9 *	71.5
8	65.4	76.0	76.5 *
9	23.6	22.9	24.1 *
10	60.4	57.6	71.0 *
11	45.6	54.5 *	44.6
12	61.8	73.0 *	69.3
13	63.5 *	59.4	61.8
14	75.3	75.0	76.0 *
15	74.9	73.7	75.8 *
16	42.6	51.0 *	44.1
17	63.7	67.5	68.5 *
18	42.5	43.3 *	33.9
19	67.8	70.0	75.4 *
20	52.7	56.4 *	52.6

 Table 2. The Intersection Over Union (IoU) scores of three methods.

\* Means the best score in a category.

Table 3. The IoU scores of four methods.

Category	FCN-32s	FCN-16s	FCN-8s	Ours
0	88.611	89.704	89.961 *	89.721
1	66.803	74.143	74.346	74.676 *
2	30.772	35.861	36.759	36.899 *
3	72.928	77.133	77.016	77.426 *
4	50.662	51.285	52.407 *	52.357
5	62.929	67.623	67.688	67.718 *
6	74.483	76.308 *	75.414	75.484
7	70.754	69.941	71.397	71.527 *
8	73.813	76.879 *	76.312	76.452
9	21.913	23.666	23.923	24.133 *
10	64.795	69.183	69.738	71.048 *
11	40.358	42.070	44.482	44.562 *
12	63.655	69.423 *	69.171	69.331
13	57.512	61.184	61.781	61.801 *
14	72.333	75.706	75.693	75.973 *
15	71.729	75.618	75.740	75.830 *
16	40.177	45.430 *	44.259	44.079
17	63.780	68.546 *	68.189	68.469
18	32.200	35.057 *	34.077	33.877
19	74.669	72.746	75.508 *	75.378
20	53.204 *	51.937	52.681	52.561

\* Means the best score in a category.

It can be shown in both Tables 2 and 3 that our method has an advantage in most of the classification indicators. Because Table 3 is consistent with the results obtained in Table 2, so only the Table 3 is used to analyze the performance of the proposed method, as follows:

The comparison of IoU between our proposed algorithm and the FCN model are shown in Figure 9. The abscissa value may range from 0~20, representing the different categories in PASCAL VOC 2012. Figure 9 is plotted by the data in Table 3, which gives all the IoU in 21 categories.



Figure 9. Segmentation accuracy of four methods.

For all of the four methods, the highest and the lowest IoU scores are observed in category 0 and category 9, respectively. When compared with the FCN-32s, FCN-16s, and FCN-8s, the proposed algorithm has a higher IoU score in categories with the superscript '\*' in Table 3.

Figure 10 shows the difference value of IoU score between our algorithm and the FCN model. Obviously, our algorithm outperforms FCN model in large majority of categories. Especially, the IoU score of FCN model are 64.79%, 69.18%, and 69.73% in category 10. But, they are 6.25%, 1.86% and 1.31% lower than our algorithm. A detailed comparison is shown as follow:



Figure 10. IoU difference value between our algorithm with FCN models.

Table 4 shows the pair-wise comparison of IoU score between our method and the three other methods. When compared with FCN-32s, 20 categories have been improved by using our proposed method. The increased category ratio is 95.24%. When compared with FCN-16s, 15 categories have been improved. The increased category ratio is 71.43%. As compared with FCN-8s, 16 categories have been improved. The increased category ratio is 76.16%. The comparison proves our method can achieve a good performance.

	Con	npare 1	Compare 2		Compare 3	
Method	Ours	FCN-32s	Ours	FCN-16s	Ours	FCN-8s
Advantage Category Increased Ratio	20 95.24%	1 % (20/21)	15 71.43%	6 % (15/21)	16 76.19%	5 6 (16/21)

Table 4. Pair-wise comparison of IoU score.

As shown in Table 5, our algorithm achieves mIoU score at 62.8% which outperforms others. Our algorithm achieves PA score at 77.14%, which is 3.86%, 1.41%, and 1.28% higher than that of FCN-32s, FCN-16s, and FCN-8s, individually.

Table 5. The mean intersection over union (mIoU) and pixel accuracy (PA) scores of all the methods.

Metric	mIoU	PA
FCN-32s	59.4	73.27
FCN-16s	62.3	75.73
FCN-8s	62.6	75.85
Ours	62.8 *	77.14 *
SegNet	59.9	-
UNIST_GDN_FCN	62.2	-

\* Means the best score in a metric.

The scores of mIoU show that the proposed algorithm has an average improvement in the classification accuracy. The scores of PA indicate that our algorithm has a better fit for the object contour. The enhancement of these indicators has benefited from the application of BSLIC and the four criteria.

### 5. Conclusions

In this paper, an improved image semantic segmentation algorithm is proposed. The BSLIC superpixel edge information is combined with the original FCN model, where superpixel semantic annotation is applied to the combination process. The superpixel semantic annotation is abiding by four criteria, so that the detailed structures of an object can be kept to some extent. The four criteria cover every situation of superpixel semantic annotation process and give each one a solution to acquire a clearer edge of the target. The algorithm not only inherits the high-level feature extraction ability of the FCN model, but it also takes full advantages of the low-level features that are extracted by BSLIC. Compared with the FCN semantic segmentation model, the pixel accuracy of the proposed algorithm on PASCAL VOC 2012 dataset is 77.14%, which is 3.86%, 1.41%, and 1.28% higher than that of FCN-32s, FCN-16s, and FCN-8s, respectively. The experiment results prove that our method is available and effective. The proposed method combines FCN with BSLIC by using the four criteria, and the introduction of other modules might provide optimized solution to extent the application.

**Author Contributions:** Wei Zhao and Haodi Zhang conceived and designed the experiments; Haodi Zhang performed the experiments; Hai Wang and Yujin Yan analyzed the data; Yi Fu contributed analysis tools; and Yujin Yan and Hai Wang wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

# References

- 1. Feldman, J.A.; Yakimovsky, Y. Decision theory and artificial intelligence: I. A semantics-based region analyzer. *Artif. Intell.* **1974**, *5*, 349–371.
- Yang, L.; Meer, P.; Foran, D.J. Multiple class segmentation using a unified framework over mean-shift patches. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, CVPR'07, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–8.
- Arbeláez, P.; Pont-Tuset, J.; Barron, J.T.; Marques, F.; Malik, J. Multiscale combinatorial grouping. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 328–335.
- Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
- 5. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic image segmentation with deep convolutional nets and fully connected crfs. *Comput. Sci.* **2014**, 357–361.
- Noh, H.; Hong, S.; Han, B. Learning deconvolution network for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Los Alamitos, CA, USA, 7–13 December 2015; pp. 1520–1528.
- Mostajabi, M.; Yadollahpour, P.; Shakhnarovich, G. Feedforward semantic segmentation with zoom-out features. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3376–3385.
- Chen, L.-C.; Yang, Y.; Wang, J.; Xu, W.; Yuille, A.L. Attention to scale: Scale-aware semantic image segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 27–30 June 2016; pp. 3640–3649.
- 9. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *arXiv* **2016**, arXiv:1606.00915.
- Taskar, B.; Abbeel, P.; Koller, D. Discriminative probabilistic models for relational data. In Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence, Edmonton, AB, Canada, 1–4 August 2002; pp. 485–492.
- Lafferty, J.D.; Mccallum, A.; Pereira, F.C.N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the Eighteenth International Conference on Machine Learning, Williamstown, MA, USA, 28 June–1 July 2001; pp. 282–289.
- 12. Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. Labelme: A database and web-based tool for image annotation. *Int. J. Comput. Vis.* **2008**, 77, 157–173. [CrossRef]
- 13. Carreira, J.; Li, F.; Sminchisescu, C. Object recognition by sequential figure-ground ranking. *Int. J. Comput. Vis.* **2012**, *98*, 243–262. [CrossRef]
- 14. Uijlings, J.R.; Van De Sande, K.E.; Gevers, T.; Smeulders, A.W. Selective search for object recognition. *Int. J. Comput. Vis.* **2013**, *104*, 154–171. [CrossRef]
- 15. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
- 16. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [CrossRef] [PubMed]
- 17. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *39*, 91–99. [CrossRef] [PubMed]
- 18. Carreira, J.; Sminchisescu, C. Cpmc: Automatic object segmentation using constrained parametric min-cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 1312–1328. [CrossRef] [PubMed]
- Liu, S.; Qi, X.; Shi, J.; Zhang, H.; Jia, J. Multi-scale patch aggregation (mpa) for simultaneous detection and segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 27–30 June 2016; pp. 3141–3149.
- 20. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, 1097–1105. [CrossRef]
- 21. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* 2014, arXiv:1409.1556.

- 22. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. *arXiv* **2015**, arXiv:1409.4842.
- 23. Ren, X.; Malik, J. Learning a classification model for segmentation. In Proceedings of the Ninth IEEE International Conference on Computer Vision, Nice, France, 13–16 October 2003; p. 10.
- 24. Levinshtein, A.; Stere, A.; Kutulakos, K.N.; Fleet, D.J.; Dickinson, S.J.; Siddiqi, K. Turbopixels: Fast superpixels using geometric flows. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 2290–2297. [CrossRef] [PubMed]
- 25. Zhang, Y.; Hartley, R.; Mashford, J.; Burn, S. Superpixels via pseudo-boolean optimization. In Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, 6–13 November 2011; pp. 1387–1394.
- Van den Bergh, M.; Boix, X.; Roig, G.; de Capitani, B.; Van Gool, L. Seeds: Superpixels extracted via energy-driven sampling. In Proceedings of the European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; pp. 13–26.
- 27. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282. [CrossRef] [PubMed]
- 28. Wang, H.; Peng, X.; Xiao, X.; Liu, Y. Bslic: Slic superpixels based on boundary term. *Symmetry* **2017**, *9*, 31. [CrossRef]
- 29. Wang, H.; Xiao, X.; Peng, X.; Liu, Y.; Zhao, W. Improved image denoising algorithm based on superpixel clustering and sparse representation. *Appl. Sci.* **2017**, *7*, 436. [CrossRef]
- Hariharan, B.; Arbeláez, P.; Bourdev, L.; Maji, S.; Malik, J. Semantic contours from inverse detectors. In Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, 6–13 November 2011; pp. 991–998.
- 31. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Garcia-Rodriguez, J. A review on deep learning techniques applied to semantic segmentation. *arXiv* **2017**, arXiv:1704.06857.
- 32. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for scene segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]
- 33. Nekrasov, V.; Ju, J.; Choi, J. Global deconvolutional networks for semantic segmentation. *arXiv* 2016, arXiv:1602.03930.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).