A Componentwise Approach to Weakly Supervised Semantic Segmentation Using Dual-Feedback Network

Zhengqiang Zhang, Qinmu Peng, Sichao Fu, Wenjie Wang, Yiu-Ming Cheung, *Fellow, IEEE*, Yue Zhao, Shujian Yu, *Member, IEEE*, and Xinge You, *Senior Member, IEEE*

***Abstract*— Recent weakly supervised semantic segmentation methods generate pseudolabels to recover the lost position information in weak labels for training the segmentation network. Unfortunately, those pseudolabels often contain mislabeled regions and inaccurate boundaries due to the incomplete recovery of position information. It turns out that the result of semantic segmentation becomes determinate to a certain degree. In this article, we decompose the position information into two components: high-level semantic information and low-level physical information, and develop a componentwise approach to recover each component independently. Specifically, we propose a simple yet effective pseudolabels updating mechanism to iteratively correct mislabeled regions inside objects to precisely refine high-level semantic information. To reconstruct low-level physical information, we utilize a customized superpixel-based random walk mechanism to trim the boundaries. Finally, we design a novel network architecture, namely, a dual-feedback network (DFN), to integrate the two mechanisms into a unified model. Experiments on benchmark datasets show that DFN outperforms the existing state-of-the-art methods in terms of intersection-over-union (mIoU).**

Manuscript received 22 August 2020; revised 10 June 2021 and 6 December 2021; accepted 15 December 2021. Date of publication 4 February 2022; date of current version 6 October 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 61772220 and Grant 62101179, in part by the Key Program for International S&T Cooperation Projects of China under Grant 2016YFE0121200, in part by the Special Projects for Technology Innovation of Hubei Province under Grant 2018ACA135, in part by the Key Science and Technology Innovation Program of Hubei Province under Grant 2017AAA017, in part by the Key Research and Development Program of Hubei Province under Grant 2020BAB027, in part by the Natural Science Foundation of Hubei Province under Grant 2021CFB332, in part by fund from the Science, Technology and Innovation Commission of Shenzhen Municipality under Grant JSGG20180507182030600, and in part by the School Project of Hubei University under Grant 202011903000002. *(Corresponding author: Qinmu Peng.)*

Zhengqiang Zhang, Qinmu Peng, Sichao Fu, Wenjie Wang, and Xinge You are with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430079, China (e-mail: xtudbxk@gmail.com; pengqinmu@hust.edu.cn; fusichao_upc@163.com; wangwj54@hust.edu.cn; youxg@mail.hust.edu.cn).

Yiu-Ming Cheung is with the Department of Computer Science, Hong Kong Baptist University, Hong Kong (e-mail: ymc@comp.hkbu.edu.hk).

Yue Zhao is with the School of Computer Science and Information Engineering, Hubei University, Wuhan 430068, China (e-mail: zhaoyue@hubu.edu.cn).

Shujian Yu is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: yusjlcy9011@ufl.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2022.3144194.

Digital Object Identifier 10.1109/TNNLS.2022.3144194

Index Terms— **Componentwise approach, deep learning, dual-feedback network (DFN), weakly supervised semantic segmentation.**

I. Introduction

SEMANTIC segmentation of an image [1], [2] refers to the task of assigning each pixel a categorical label (e.g., motorcycle or person). Owing to the rapid development of deep learning, tremendous progress has been made for fully annotated semantic segmentation. Some examples include fully convolutional network (FCN) [3], DeepLab [4], GANet [5], and SegGAN [6]. These methods assume that the pixel-level labels are available immediately upon request. However, this assumption is over-optimistic because the annotations of numerous data are laborious. As a result, interactive segmentation [7]–[11] and weakly supervised semantic segmentation, which only requires a few weak labels, such as bounding box [12]–[14], scribble [15], [16], points [17], and tags [18]–[22], have attracted increasing attention.

This article focuses on weakly supervised semantic segmentation, in which only image-level tags without any position information for the least labeling cost, as shown in Fig. 1(a). Under this circumstance, as reported in [23] and [24], the absence of position information of weak labels prevents segmentation network learning from these labels directly, which makes the tag-supervised semantic segmentation problem ill-conditioned. Consequently, how to recover the lost position information becomes a pivotal issue. In the literature, Zhang *et al.* [25] proposed the decoupled spatial neural attention (DSNA) for weakly supervised semantic segmentation, which can simultaneously utilize the object regions and localize the discriminative parts to generate high-quality pseudoannotations. Zhou *et al.* [26] proposed the image-level supervision-based watershed algorithm to solve the problem of image semantic segmentation that lacks fully supervised segmentation labels. Shimoda and Yanai [27] proposed the CNN-based class-specific saliency maps and fully connected conditional random field (CRF)-based weakly supervised semantic segmentation method to reduce its high costs of pixel-wise annotated image datasets. Furthermore, Shen *et al.* [8] proposed a knapsack constraint approximately based maximizing quadratic submodular energy method utilizing dynamic programming for motion clustering and image segmentation. In addition, Shen *et al.* [10] introduced a general higher order binary energy minimization function to solve the problem of image segmentation. In addition, Dong *et al.* [11] proposed a global and local energy-based interactive co-segmentation method. Despite the substantial progress made by these methods, they do not consider the differences between high-level semantic information and low-level physical information. As a result, the recovered position information is incomplete and the generated pseudolabels still contain incorrectly labeled regions and inaccurate boundaries, as shown in Fig. 1(d).

In this article, we will decompose the position information into two components (i.e., high-level semantic information and low-level physical information) and develop a componentwise approach to recovering two kinds of information individually [see Fig. 1(b)]. High-level semantic information is the information that describes a semantic structure for a specific category. Examples of high-level semantic information include the human body structure, vehicle structure, and so on. Given the position of a human head, it is easy to infer the region of the neck using human body structure. By contrast, low-level physical information is at the other end of the spectrum. Although high-level semantic information affects the logical structure of categories, low-level physical information describes the physical structure in a single image. For example, given a region of one category, it is probably that regions with similar textures or colors belong to the same category.

Accordingly, we propose two different mechanisms to recover the high-level semantic information and low-level physical information upon original pseudolabels, respectively. The first mechanism, named pseudolabels updating mechanism, aims to correct the mislabeled regions inside objects to recover finer logical structures for categories. It uses a simple yet effective weighted updating process, which routes segmentation network output back to update the original pseudolabels. It is worth noting that deep seed region growing (DSRG) [20] and mining common object features (MCOFs) [22] both also update original pseudolabels. However, they focus on expanding discriminative regions without correcting the mislabeled regions in pseudolabels, which constrains the upper bound of segmentation performance. By contrast, the second mechanism, named customized superpixel-based random walk mechanism, is utilized to trim the boundaries in original pseudolabels to fit better with the physical structure of each image. This mechanism utilizes superpixels as the base unit of pseudolabels due to their flexible shape and size. To overcome the over-segmented problem, we propose a customized random walk process, which makes use of a novel relationship matrix and additional threshold functions to generate robust and confident pseudolabels with the help of network outputs. Although MCOF [22] and superpixel pooling network (SPN) [28] also suggest using superpixels for weakly supervised semantic segmentation, they ignore the over-segmented phenomenon, thus resulting in lots of redundant boundaries in pseudolabels. Furthermore, Ahn and Kwak [29] have introduced a random walk process to generate pseudolabels as well. However, they have to introduce an extra network to learn the semantic affinity matrix and do not notice the unequal confidence problem (i.e., regions with different probability distribution have different confidences), which results in numerous parameters and worse pseudolabels. In addition, to unify these two mechanisms into a joint framework, we will introduce two feedback chains to Deeplab-LargeFOV network [30], as shown in Fig. 1(c), which results in the dual-feedback network (DFN).

To sum up, our contributions are twofold.

1) We propose a componentwise approach and interpret the position information as two distinct categories: high-level semantic information and low-level physical information. Accordingly, the pseudolabels updating mechanism and the customized superpixel-based random walk mechanism are proposed to compensate for the lack of one type of position information.
2) DFN is proposed to implement the above-mentioned two mechanisms. The first feedback chain uses a simple yet effective weighted updating process, which will correct the mislabeled regions inside objects in original pseudolabels, to recover finer logical structures for categories. The second feedback chain uses a customized superpixel-based random walk process, with the help of the novel relationship matrix and additional threshold functions, to trim the boundaries in original pseudolabels to fit better with physical structures in each image.

II. Related Work

Compared with fully supervised semantic segmentation, the main issue for weakly supervised semantic segmentation under tags supervision is the lack of position information that describes the positions of the tagged objects in an image. In order to recover such kind of information, a number of efforts have been made in recent years, which can be divided into two categories.

The first category uses a specialized loss function or network structure to add prior knowledge of position information into the framework. For example, to locate pixels of different objects, Pathak *et al.* [31] initiated a multi-instance loss (MIL) function to restrict the network output and only those pixels which are important for locating are considered. Latter, Kolesnikov and Lampert [19] proposed the global weighted rank pooling (GWRP) method to constrain all pixels and put more weights on pixels, which are easier to classify. Meanwhile, Pathak *et al.* [32] added lots of linear constraints, such as size constraint for each tag, on the output space of the network, thereby leading to a more complete recovery of position information. To exploit the full object for each category, Wei *et al.* [33] applied the adversarial erasing mechanism to mine nondiscriminative regions iteratively in erased images. Although considerable progress has been made recently along improving the loss function or network structure, this kind of method has a dynamic optimization direction in the training procedure, which makes it hard to reconstruct complete and accurate position information.

Different from the first category, the methods of the second category generate pseudolabels to recover the missing position information. The generated pseudolabels are relatively fixed and lead to a steady training procedure. Kolesnikov and Lampert [19] obtained the sparse pseudolabels by locating the foreground regions with the class activation map (CAM)

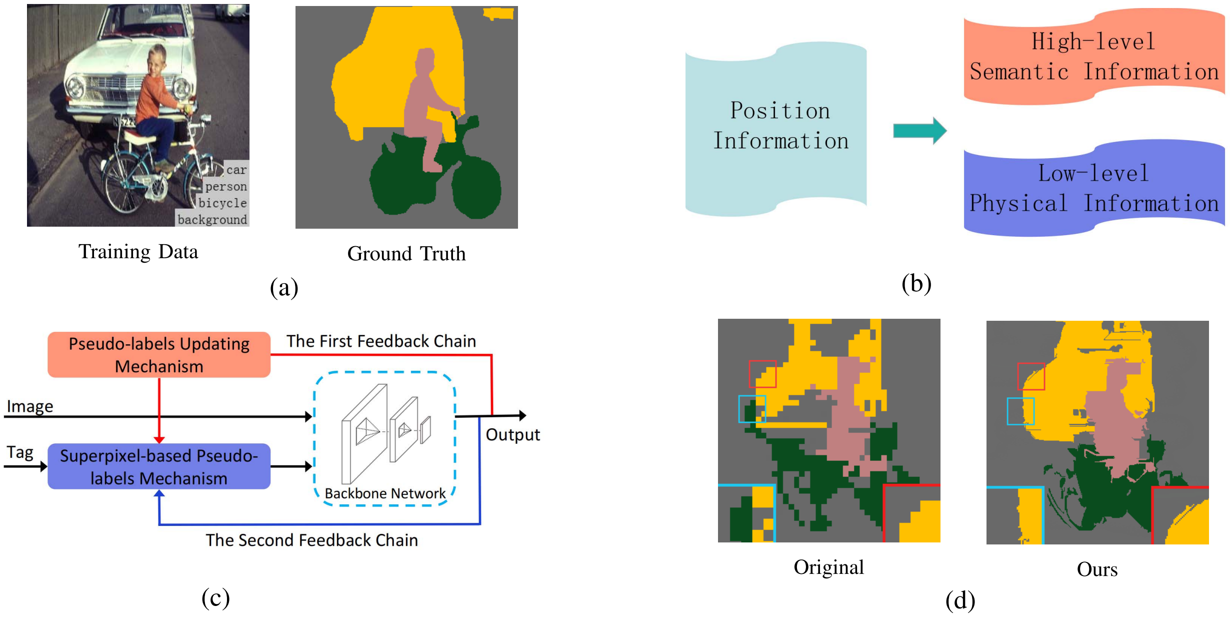

car
person
bicycle
background
Training Data
Ground Truth
(a)
Position Information
High-level Semantic Information
Low-level Physical Information
(b)
Pseudo-labels Updating Mechanism
The First Feedback Chain
Image
Tag
Superpixel-based Pseudo-labels Mechanism
Backbone Network
Output
The Second Feedback Chain
(c)
Original
Ours
(d)

Fig. 1. (a) Example of training data and its ground truth. (b) Illustration of the componentwise approach. (c) Pipeline of our DFN. The two feedback chains implement two different mechanisms to rebuild two kinds of position information. (d) Generated pseudolabels of previous methods and ours. The original pseudolabels contain incorrectly labeled regions, such as the region labeled as bicycle on the right of the car, as shown in the blue box, and have worse boundaries, such as the inaccurate segmentation boundaries between car and background, as shown in the red box.

method [34] and the background regions with [35]. Then, Shimoda and Yanai [36] utilized the distinct class saliency maps (DCSM) and Huang *et al.* [20] utilized the discriminative regional feature integration (DRFI) [37] to generate more accurate pseudolabels. For the pseudolabels of complex images which contain a number of objects, Wei *et al.* [24] proposed a simple-to-complex (STC) mechanism to produce pseudolabels progressively. Recently, Huang *et al.* [20] have utilized the classical seeded region growing method to expand regions gradually and generate dense pseudolabels. In addition, Wang *et al.* [22] have introduced a new region classification network to update static pseudolabels and obtain the improved pseudolabels iteratively during the training phase. Despite substantial progress made by these methods, they do not consider the differences between high-level semantic information and low-level physical information. As a result, the generated pseudolabels still contain incorrectly labeled regions and inaccurate boundaries, as shown in Fig. 1(d).

Following our componentwise approach, the proposed DFN utilizes two feedback chains, one uses pseudolabels updating mechanism to correct mislabeled regions hidden in original pseudolabels and the other uses superpixel-based pseudolabels mechanism to obtain accurate and concise boundaries in pseudolabels, to reconstruct the complete position information. It is worth noting that DSRG [20] also updates original pseudolabels. However, they focus on expanding discriminative regions and do not correct the mislabeled regions in pseudolabels, which constrain the upper bound of segmentation performance. On the other hand, MCOF [22] also suggests using superpixels in pseudolabels. But they do not integrate low-level physical information among superpixels and have to introduce an extra network to classify superpixels, which results in numerous redundant boundaries inner objects. Moreover, experiments demonstrate that our network outperforms both DSRG and MCOF on PASCAL VOC 2012 segmentation set and the COCO dataset. Our proposed DFN utilizes random walk techniques to trim the boundaries. In recent years, random walk techniques have been widely used, For example, Grady [38] acquired a similar probability between labeled pixels and unlabeled pixels via random walks for multilabel and interactive images segmentation. Dong *et al.* [39] proposed a label prior-based sub-Markov random walk on a graph to add the auxiliary nodes, and then applied for the image segmentation. Shen *et al.* [40] first utilized the lazy random walks to acquire the probabilities of each pixel and then proposed a commute time and the texture measurement-based energy function for superpixel segmentation.

III. Proposed Method

A. Problem Definition

Tag-supervised semantic segmentation problem aims to train a semantic segmentation model only using natural images with their corresponding tags. Specifically, there exist two totally different datasets: a training set $D_{\text{train}} = \{(X_i, T_i)\}_{i=1}^{N_{\text{train}}}$ and a separated testing set $D_{\text{test}} = \{(X_k, Y_k)\}_{k=1}^{N_{\text{test}}}$, where X represents the nature image, while T and Y are the corresponding image tags and segmentation mask. The problem requires us to train the deep neural network only on the training set D_{train} and evaluate the segmentation performance on the testing set D_{test}. Obviously, due to the missing semantic segmentation mask Y_i in D_{train}, it is hard to train a semantic segmentation model directly on it. Recently, researchers have found that the key to this problem is to recover the missing position information to reconstruct the missing Y_i. To this end, this article first generates the pseudolabels Y_i^{pseudo} given the training pairs

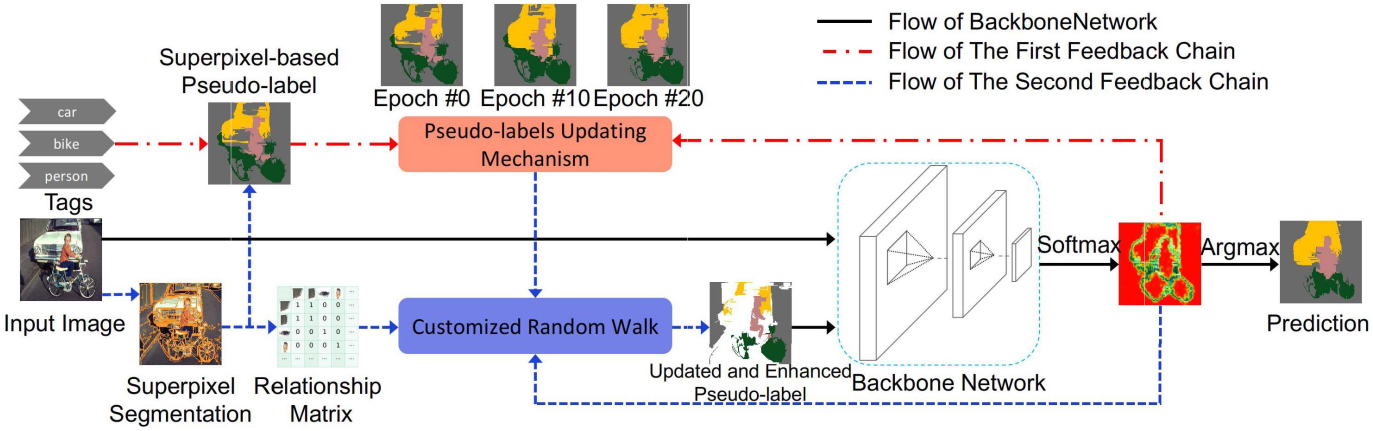


Fig. 2. Schematic sketch of the proposed DFN with two feedback chains. The first feedback chain (red arrow) dynamically updates the pseudolabels with the aid of network output. The second feedback chain (blue arrows) merges the relationship matrix, the network output, and the updated pseudolabels using a customized random walk process. The input to the backbone network includes both the training image and the updated and enhanced pseudolabels.

(X_i, T_i). Then, Y_i^{pseudo} is used to supervise the neural network to learn the capability of semantic segmentation. In testing, the trained model outputs $\hat{Y}_k$ for each X_k, and the performance can be evaluated by the overlap between Y_k and $\hat{Y}_k$.

B. The Proposed Approach

We decompose the improvement of pseudolabels into two different mechanisms. One is the pseudolabels updating mechanism, which updates the pseudolabels with the aid of the output of the backbone network to recover the high-level semantic information. The other is the customized superpixel-based random walk mechanism, which enhances the pseudolabels with superpixel segmentation and customized random walk process to recover the low-level physical information. Specifically, analogous to [19], we use CAM [34] and object detection map generated by DRFI approach [37] to generate original pseudolabels. The CAM for each class is used to discriminate which pixel belongs to the foreground, and the object detection map is used to discriminate which pixel belongs to the background. Then, as shown in Fig. 2, our DFN integrates two feedback chains into a backbone network from Deeplab [30] to update and enhance the original pseudolabels. The first feedback chain, which implements the pseudolabels updating mechanism, utilizes a simple weighted updating operation to correct mislabeled regions. The second feedback chain, which implements the customized superpixel-based random walk mechanism, introduces the superpixel segmentation and customized random walk process to trim the boundaries in original pseudolabels. Finally, we obtain the updated and enhanced pseudolabels and utilize them to train the backbone network under the seed loss function and the boundary loss function [19]. The formulas of each loss function for single image are:

$$L_{\text{seed}} = -\frac{\sum_{i=1}^{N}\sum_{k=1}^{M_i} \log\ p_{i,k}}{\sum_{i=1}^{N} M_i} - \frac{\sum_{k=1}^{M_0} \log\ p_{0,k}}{M_0} \tag{1}$$

where N denotes the number of tags in each image, M_i denotes the number of pixels of tag i (0 represents the background), and $p_{i,k}$ is the probability that pixel k is classified into tag i

$$L_{\text{constrain}} = -\frac{\sum_{i=1}^{N}\sum_{k=1}^{M} q_{i,k} \log\ p_{i,k}}{NM} \tag{2}$$

where N and M denote the number of tags in dataset and the number of pixels of single image, respectively, and $q_{i,k}$ represents the CRF-processed [41] probability distribution.

C. Recovery of High-Level Semantic Information

With CAM and DRFI techniques, we generate the original pseudolabels which mark the discriminative regions for each tag in the corresponding image. Those discriminative regions usually implicitly contain the logical structure of a special category, e.g., person. Unfortunately, original pseudolabels still contain some errors hidden in labeled regions, as shown in Fig. 1(d). We have to correct those errors for the purpose of reconstructing finer high-level semantic information. As stated in Zhang *et al.* [14], the iterative learning strategy has been widely used under the weakly supervised framework. Therefore, we also utilize a simple yet effective pseudolabels updating mechanism to correct mislabeled regions. Specifically, we utilize an independent feedback chain to implement it with the help of the backbone network. The backbone network is trained with original pseudolabels and learns a better logical structure for each category across images. Then, the feedback chain routes its output back into original pseudolabels to correct the mislabeled regions. The pseudolabels updating mechanism repeats the procedure several iterations. In each updating process, we update pseudolabels with the network output using the following formula:

$$P^i = (1 - w) \times P^{i-d} + w \times N_{\text{out}}^i \tag{3}$$

where P denotes pseudolabels, N_{out}^i is the network output, i denotes the ith iteration, d is the updating interval, and $w \in [0, 1]$ is a weighting factor that determines the update rate.

One should note that there are three key points behind (3). First, there exist many training epochs before updating the original pseudolabels to ensure the segmentation network reaches a good point for extracting the high-level semantic

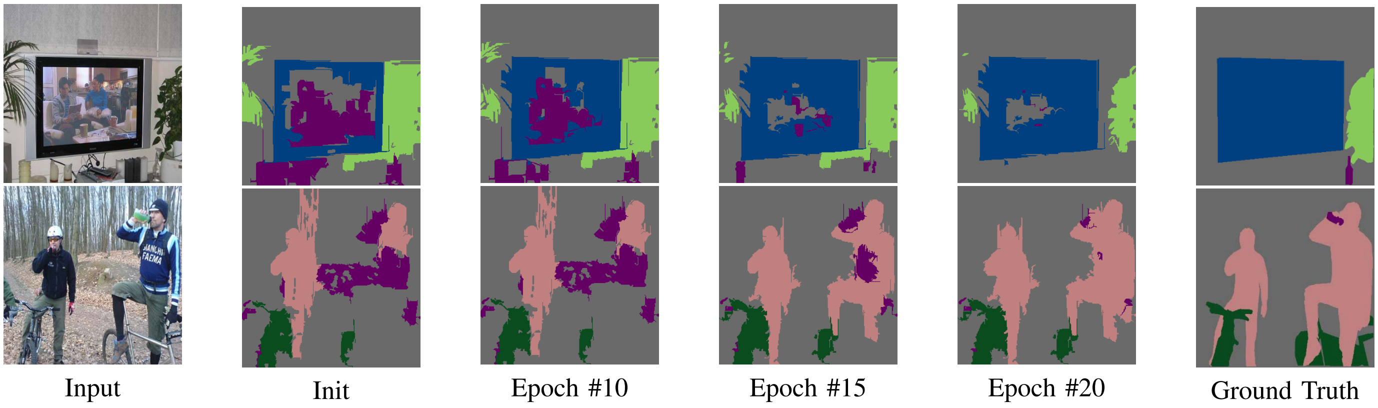


Fig. 3. Examples of the updated pseudolabels generated by the first feedback chain in several epochs during the training phase.

information. Second, in the interval of two subsequent updating operations, the network is trained independently with a few epochs to make full use of updated pseudolabels. Moreover, different from Seed, Expand and Constrain (SEC) [19], each base unit in pseudolabels is characterized by a vector with its element corresponding to the probability to each category, which makes it convenient to update. As can be seen in Fig. 3, with the increase of the number of iterations, initial errors are corrected and the logical structures get closer to the ground truth, which means high-level semantic information is reconstructed progressively. The overall procedure is summarized as Algorithm 1.

Algorithm 1 Pseudolabels Updating Mechanism

```
Input:      origin pseudo-labels $P^0$, the update weight $w$,
            the updating interval iteration $d$,
            the iteration to start updating $t_{start}$,
            the total training iteration $T$.
Procedure:
            1. for $i$-th training iteration in range from 1 to $T$
            2.   get $N_{out}^i$ from network
            3.   updating network parameters with $P^i$
            4.   if $i$ is bigger than $t_{start}$ and is evenly divided by $d$
            5.     $P^i = (1 - w) \times P^{i-d} + w \times N_{out}^i$
            6.   else
            7.     $P^i = P^{i-1}$
Output:     the trained network
```

D. Recovery of Low-Level Physical Information

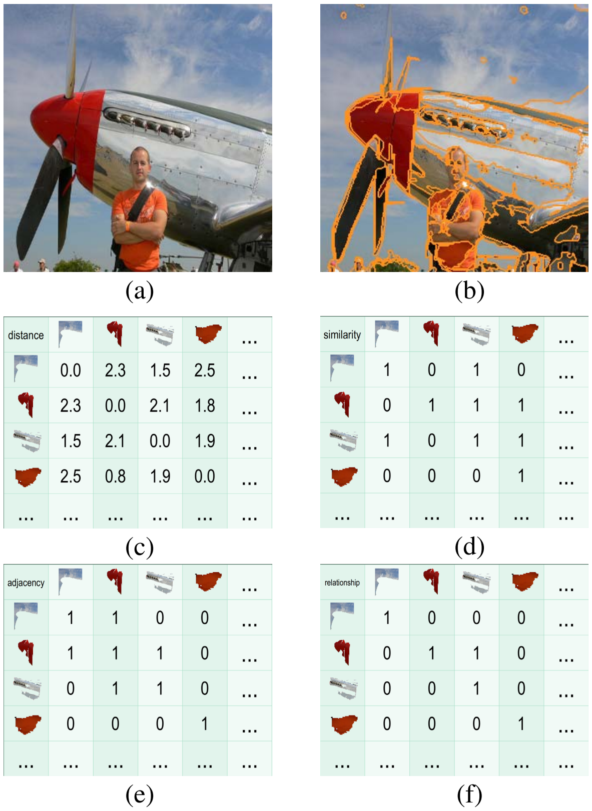
To recover missing low-level physical information, we set out to trim inaccurate and redundant boundaries in original pseudolabels. Therefore, we propose a customized superpixel-based random walk mechanism, which migrates the boundaries from superpixel segmentation into origin pseudolabels.

1) Superpixel Based Pseudolabels: Owing to the low spatial resolution of deep convolutional network output, previous methods always utilize square pixels blocks as the basic unit to store pseudolabels for each image. Each block is characterized by a scalar that represents the tag of a box of corresponding pixels whose size is 8 × 8 in an input image. Unfortunately, due to the fixed shape and size of the square pixels block, it is hard to learn the details of boundaries in the backpropagation process as mentioned in [42] and rebuild the missing low-level physical information.

To solve this problem, this article proposes to use superpixels, which have varying sizes and sophisticated boundaries, to replace square pixels blocks as the basic unit of pseudolabels. The superpixel-based pseudolabels have the same spatial resolution as the input image, which makes the network keep the details during the backpropagation process. To achieve this goal, we have to segment input image into superpixels with normal superpixel segmentation methods, such as felzenszwalb [43], as shown in Fig. 4(b), and then convert the old pseudolabels whose basic unit is square pixels blocks into the new pseudolabels using superpixels. As mentioned earlier, each superpixel in the new pseudolabel uses a vector of classification probability to describe the labels. In detail, we convert the scalar in each pixel into the one-hot classification probability. Then, each superpixel gets a vector averages among all the one-hot probabilities of pixels belonging to the corresponding superpixel.

2) Customized Random Walk Process: Unfortunately, there still exists an over-segmentation phenomenon for all superpixel segmentation methods, as mentioned in [44]. This phenomenon refers to that an image is segmented into a number of small superpixels, which introduces numerous unnecessary boundaries and cuts off the fusion of low-level physical information inner one object. To prevent the over-segmentation phenomenon, we introduce a relationship matrix to describe the similarity and adjacency between any two superpixels inner one image and use it to merge the low-level physical information by our customized random walk process.

a) Relationship matrix: The construction of our relationship matrix consists of four steps. First, we generate Euclidean distance matrix [see Fig. 4(c)] for superpixels using zoom-out features [45]. However, unlike them, we only use the first two convolutional layers in VGG-16 [46] to obtain low-level physical features. Nevertheless, owing to the diversity of objects under complex backgrounds in each image,



(a) (b)

distance					...
	0.0	2.3	1.5	2.5	...
	2.3	0.0	2.1	1.8	...
	1.5	2.1	0.0	1.9	...
	2.5	0.8	1.9	0.0	...
...	...	...	...	...	...

(c)

similarity					...
	1	0	1	0	...
	0	1	1	1	...
	1	0	1	1	...
	0	0	0	1	...
...	...	...	...	...	...

(d)

adjacency					...
	1	1	0	0	...
	1	1	1	0	...
	0	1	1	0	...
	0	0	0	1	...
...	...	...	...	...	...

(e)

relationship					...
	1	0	0	0	...
	0	1	1	0	...
	0	0	1	0	...
	0	0	0	1	...
...	...	...	...	...	...

(f)

Fig. 4. Top row orderly shows a training image and its corresponding superpixel segmentation result given by [43]. The next two rows show the Euclidean distance matrix, the relative similarity matrix, the distance matrix, and the relationship matrix. (a) Input. (b) Superpixel. (c) Distance Matrix. (d) Similarity Matrix. (e) Adjacency Matrix. (f) Relationship Matrix.

the Euclidean distance with absolute values cannot precisely reflect the similarity between any two superpixels. Therefore, we transform the Euclidean matrix to a simple relative similarity matrix (denoted M_{siml}), as shown in Fig. 4(d). Specifically, we pick the ten smallest values in each row of the Euclidean matrix and set their values to 1, indicating that they are similar. We then set the values of the remaining elements in each row of the Euclidean matrix to 0, suggesting that they are not similar. Next, we construct an adjacency matrix (denoted M_{adj}) for superpixels to incorporate their similarity in geometric space. If two superpixels are neighboring, the corresponding element in the adjacency matrix is 1, otherwise 0 [see Fig. 4(e)]. We finally obtain a relationship matrix (denoted M_{rel}) with the following formula [see Fig. 4(f)]:

$$M_{\mathrm{rel}} = M_{\mathrm{siml}} \odot M_{\mathrm{adj}} \tag{4}$$

where $\odot$ denotes entrywise product.

b) Unequal confidence problem: After recovering the low-level physical information characterized by the relationship matrix built upon superpixels, we apply the random walk process to incorporate low-level physical information into network training by using the relationship matrix as the transition probability matrix and superpixel-based pseudolabels as the initial state. However, a serious problem, named unequal confidence problem, arises when directly adopting the normal random walk process. Specifically, each superpixel in pseudolabels is characterized by a classification probability vector and it means that if the value of the maximum element in the vector (of a superpixel) is small, we have a high classification uncertainty to assign this superpixel to its corresponding category. Moreover, the superpixels with high classification uncertainty will mislead the direction of pseudolabels changing in each iteration of the random walk process.

Therefore, we propose two simple methods to improve the normal random walk process. At first, we filter the unconfident superpixels out of pseudolabels during the random walk process, which ensures all superpixels inputted to the random walk process have high confidence. In essence, we use a simple threshold function [denoted $\mathrm{Th}_1(\cdot)$] to achieve this goal and it is coupled with two hyperparameters α_{fg} and α_{bg} for foreground and background in pseudolabels, respectively, due to the different complexities to locate related regions. For example, if the maximum element of one superpixel belonging to the foreground in pseudolabels is less than α_{fg}, we dropout this superpixel in the following random walk process. Next, to ensure the high confidence of new superpixel outputted from the random walk process, we route network output back to confirm it along the second feedback chain. Specifically, network output is resized into the same size as the input image using bilinear interpolation, and the new superpixel obtains its corresponding probability averaged all the probabilities of pixels belonging to this superpixel. Then, we also use another threshold function [denoted $\mathrm{Th}_2(\cdot)$], coupled with hyperparameters β_{fg} and β_{bg} for foreground and background, to filter out unconfident superpixels on the generated new pseudolabels.

In summary, the second feedback chain performs an improved random walk process to suppress the over-segmentation phenomenon. The input to this customized random walk includes the relationship matrix which describes the similarity and adjacency among superpixels, pseudolabels whose basic unit is superpixels, and the output of backbone network. By using the superpixel-based pseudolabels as the initial state and the relationship matrix as the transition probability matrix, the random walk process generates expanded and alternative superpixels with "$P \times M_{\mathrm{rel}}$," where "×" denotes matrix multiplication. However, to solve the unequal confidence problem, we use two threshold functions [$\mathrm{Th}_1(\cdot)$ and $\mathrm{Th}_2(\cdot)$] to filter those superpixels out with aid of network output N_{out}. Therefore, the result $\hat{P}$ of our one-step customized random walk process is given by

$$\hat{P} = \mathrm{Th}_1(P) \times M_{\mathrm{rel}} \odot \mathrm{Th}_2(N_{\mathrm{out}}) \tag{5}$$

where $\odot$ denotes the entrywise product. We repeat the operation of "$\times M_{\mathrm{rel}} \odot \mathrm{Th}_2(N_{\mathrm{out}})$" n times on the thresholded initial state $\mathrm{Th}_1(P)$ to obtain the n-step customized random walk result. For example, the result of two-step customized random walk is given by

$$\hat{P} = \mathrm{Th}_1(P) \times M_{\mathrm{rel}} \odot \mathrm{Th}_2(N_{\mathrm{out}}) \times M_{\mathrm{rel}} \odot \mathrm{Th}_2(N_{\mathrm{out}}). \tag{6}$$

Moreover, we present two examples of our customized random walk process in Fig. 5. By merging low-level physical

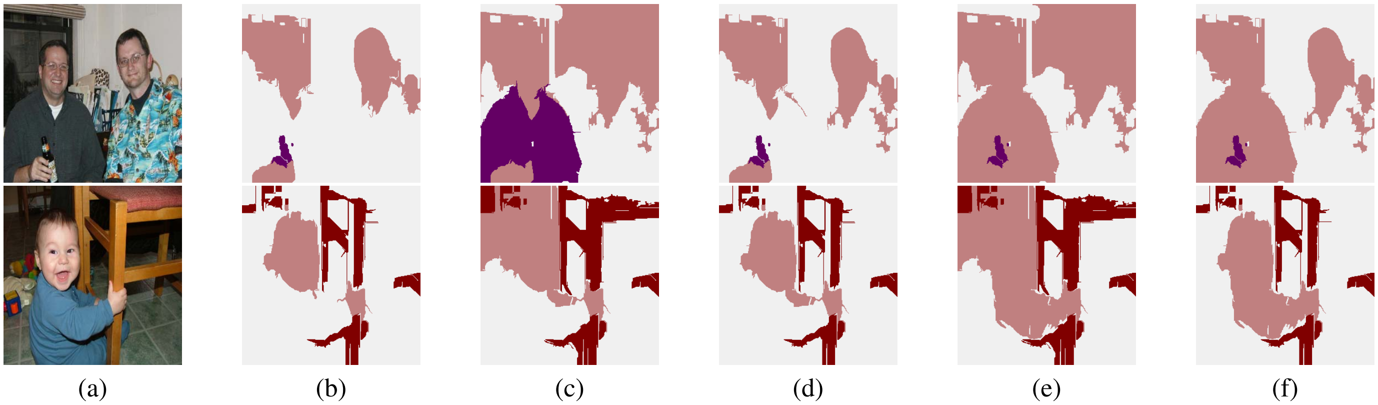


Fig. 5. Evolution of pseudolabels with our customized random walk process: (a) training images; (b) filtered superpixel-based pseudolabels; (c) one-step result of normal random walk; (d) one-step result of customized random walk; (e) two-steps result of normal random walk; (f) two-steps result of customized random walk. The normal random walk encounters unequal confidence problems and changes toward wrong direction while our customized random walk produces high confident results.

information, the enhanced pseudolabels change toward the right direction and fit well with the object boundaries.

The overall procedure of the superpixel-based random walk mechanism is summarized as Algorithm 2.

Algorithm 2 Superpixel-Based Random Walk

```
Input:      origin pixel-based pseudo-labels $P^0$,
            the superpixel segmentation $S$,
            the threshold functions $Th_1$ and $Th_2$,
            the random walk step $n$
Procedure:
            1. generate superpixel-based pseudo-labels
               $P$ using $P^0$ and $S$
            2. generate relative similarity matrix $M_{siml}$
               and adjacency matrix $M_{adj}$ by $S$
            3. generate relationship matrix:
               $M_{rel} = M_{siml} \odot M_{adj}$
            4. for $i$-th training iteration
            5.     get $N_{out}$ from network
            6.     repeate n times to enhance pseudo-
                   labels: $\hat{P} = Th_1(P) \times M_{rel}$
                   $\odot Th_2(N_{out})$
            7.     updating network parameters with $\hat{P}$
Output:     the trained network
```

E. Dual-Feedback Algorithm

In the summary, our DFN simultaneously utilizes the pseudolabels updating mechanism and the superpixel-based random walk mechanism. The entire algorithm can be described by Algorithm 3.

Algorithm 3 Dual-Feedback Algorithm

```
Input:      origin pixel-based pseudo-labels $P^0$,
            the superpixel segmentation $S$,
Procedure:
            1. generate superpixel-based pseudo-labels
               $P$ using $P^0$ and $S$
            2. for i-th training iteration in the range
               from 1 to $T$
            3. get $N_{out}$ from network
            4. run the Pseudo-labels Updating
               Mechanism to get the updated pseudo-
               labels $P_i$ using $N_{out}$
            5. run the Superpixel-based Random Walk
               Mechanism to get enhanced $\hat{P}$ using $P_i$
               and $N_{out}$
            6. updating network parameters with $\hat{P}$
Output:     the trained network
```

IV. Experiments

A. Experiment Setup

1) Dataset and Evaluation Metrics: We evaluate the proposed DFN on the PASCAL VOC 2012 segmentation benchmark dataset [47], which contains 20 foreground object classes and one background class, and the COCO dataset [48], which contains 80 foreground object classes and one background class. In detail, the segmentation part of the PASCAL VOC 2012 dataset is split into three parts: training (train, 1464 images), validation (val, 1449 images), and testing (test, 1456 images). Analogous to [20] and [22], the training set is extended with the additional images from [49], resulting in an augmented set of 10 582 images. On the other hand, the COCO dataset is split into two distinct parts: training (train, 82 784 images) and validation (val, 40 505 images). Following the common practice, we use the mean intersection-over-union (mIoU) criterion averaged on all classes to compare our method with the other approaches on val or test sets. The frequency weighted IoU (f.w. IoU) [3] and accuracy (accu) are also used to evaluate our methods under the different experiment settings. In addition, we report our results on standard val set when the ground truth segmentation masks are available. For the test set, we submit the results of our final best model to the official evaluation server.

2) Implementation Details: Following the previous methods [19], a slightly modified VGG-16 network is used for classification. Then, we adopt CAM [34] to extract pseudolabels for the foreground classes, and DRFI [37] to extract pseudolabels for the background class. For all segmentation experiments, such as SEC [19], STC (2017) [24], combining bottom-up, top-down, and smoothness cues (CBTS) (2017) [50], adversarial erasing-prohibitive segmentation learning (AE-PSL) (2017) [33], background estimation and built-in priors (BEBP) (2018) [23], watershed algorithm with image-level supervision (WAILS) (2019) [26], and MCOF [22], we use the DeepLab-LargeFOV network [30] based on VGG-16 [46] as the backbone segmentation network which is pretrained on ImageNet [51]. Moreover, we use a mini-batch of six images for SGD and an initial learning rate of $3e$-3, which is decreased by a factor of 3 every five epochs. The momentum is 0.9, the dropout rate is 0.5 and the total epochs of training is 20 at which the superpixel-based pseudolabels do not change.

In addition, for the first feedback chain, we set w to 0.2 and start to update the pseudolabels every three epochs after the tenth epoch. For the second feedback chain, we set α_{fg} to 0.90, α_{bg} to 0.90, β_{fg} to 0.75, and β_{bg} to 0.90. Moreover, we perform two steps of customized random walk. In the test phase, same to [19], the fully connected CRF [52], and the multiscale prediction [4] are applied with their default parameters.

B. Comparison With Previous Methods

1) PASCAL VOC 2012 Dataset: Tables I and II report the mIoU values of our method on PASCAL VOC 2012 val and test sets, respectively, against the previous state-of-the-art methods, namely, EM-Adapt (2015) [18], SEC (2016) [19], STC (2017) [24], CBTS (2017) [50], AE-PSL (2017) [33], BEBP (2018) [23], MCOF (2018) [22], DSRG (2018) [20], AffinityNet (2018) [29], H&S (2019) [53], DSNA (2019) [25], WAILS (2019) [26], and Easiness (2020) [27]. It can be seen that our method outperforms all compared methods in terms of mIoU value under the same experimental setting. In addition, although MCOF [22] also uses superpixels as the base unit for pseudolabels, our method introduces a customized random walk process with a robust relationship matrix to tide over the over-segmented problem and achieves a performance gain of 3.8% and 3.5% on val and test sets, respectively. The improvement is 1.0% and 0.7%, respectively, compared with DSRG [20], which only expands labeled regions in original pseudolabels and does not correct mislabeled regions directly. Compared with AffinityNet [29], our customized random walk process utilizes additional threshold functions to overcome the unequal confidence problem and gains 0.9% and 0.3% improvement on val and test sets, respectively.

2) COCO Dataset: We also conduct experiments on the COCO dataset to demonstrate the generality of our DFN. The involved methods include SEC(2016) [19], BEBP(2018) [23], DSRG(2018) [20], WAILS(2019) [26], and WSIF(2020) [54]. Most images in the COCO dataset have a more complex background and are closer to natural scenes. Moreover, there exist nearly 80k and 40k images for training and val set, respectively, which has far more data than PASCAL VOC 2012 set. Therefore, the high complexity and enormous quantity make it hard to train a weakly supervised semantic segmentation network with this dataset. Table III shows the comparison results to a few previous works in mIoU and f.w. IoU. From Table III, we can see that our DFN indeed outperforms the previous works and has high generality. In detail, our DFN reaches 26.8% and 67.6% in mIoU and f.w. IoU, respectively. The relative gains in f.w. IoU (22.2% compared with BFBP, 11.4% compared with SEC, 7.8% compared with WAILS, 1.8% compared with DSRG, and 0.4% compared with WSIF) is smaller than it in mIoU (31.4% compared to BFBP, 19.6% compared to SEC, 19.1% compared to WAILS, 3.1% compared to DSRG and 1.9% compared to WSIF), which demonstrates our method works better with the unbalanced dataset, due to f.w. IoU puts more weight on majority.

3) Different Supervision Types: We also compare our network with other methods under different types of supervisions. They are FCN [3], DeepLab [4], weakly and semi-supervised learning (WSSL) [18], BoxSup [12], random-walk (RAWK) [55], ScribbleSup [15], and What'sPoint [17]. As can be seen in Table IV, our network achieves comparable performance to other methods that require stronger supervisions, e.g., the WSSL, the RWAK, or even the fully supervised FCN. It suggests that our componentwise approach indeed reconstructs most of the missing position information. This result also suggests that there is a large performance gap between fully supervised semantic segmentation and weakly supervised semantic segmentation, especially for point supervisions.

C. Quantitative Results

The segmentation results shown in Fig. 6 corroborate our quantitative evaluations. Note that our method can generate precise segmentation results even for images containing complex backgrounds. However, our method is likely to fail when there are multiple small and dense objects on top of another larger object. Let us take the last row of Fig. 6 as an example. Our method mislabels most of the pixels in the table into the background. One possible reason is that there are a number of plates and foods on the table, such that various colors and textures of these small and dense objects make our method hard to generate a precise relationship matrix to prevent the over-segmentation phenomenon.

D. Ablation Studies

To validate the effects of different components, we perform some ablation experiments under different settings. In Table V, we summarize the performance of our network in different degrading settings. Specifically, the "baseline" indicates our baseline network without two feedback chains, which is the same as the SEC network. Similar to DSRG, due to the precise background mask generated by DRFI, the final result is better than it reported in the original paper. The "baseline + F1" indicates only integrating the first feedback chain into the baseline network. With this feedback chain, we integrate high-level semantic information from network output and correct mislabeled regions in original pseudolabels. Then, the "baseline + F2" indicates to use the second feedback chain only in the network. The customized random walk process recovers

TABLE I

COMPARISON OF DIFFERENT WEAKLY SUPERVISED SEMANTIC SEGMENTATION METHODS ON PASCAL VOC 2012 VAL SET. THE "-" DENOTES UNKNOWN VALUES THAT WERE NOT REPORTED BY THE CORRESPONDING PAPER. THE "OURS1" DENOTES THE OUR NETWORK BASED ON VGG-16 NETWORK WHILE "OURS2" DENOTES THE OUR NETWORK BASED ON RESNET-101 NETWORK

Method	bkg	plane	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	motor	person	plant	sleep	sofa	train	tv	**mIoU**
EM-Adapt	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	38.2
H & S	78.1	51.3	18.2	46.6	31.9	35.1	58.5	50.6	50.2	16.5	40.3	27.4	45.1	47.9	54.9	32.1	29.5	46.0	25.7	48.7	45.1	41.9
BEBP	79.2	60.1	20.4	50.7	41.2	46.3	62.6	49.2	62.3	13.3	49.7	38.1	58.4	49.0	57.0	48.2	27.8	55.1	29.6	54.6	26.6	46.6
STC	84.5	68.0	19.5	60.5	42.5	44.8	68.4	64.0	64.8	14.5	52.0	22.8	58.0	55.3	57.8	60.5	40.6	56.7	23.0	57.1	31.2	49.8
SEC	82.4	62.9	26.4	61.6	27.6	38.1	66.6	62.7	75.2	22.1	53.5	28.3	65.8	57.8	62.3	52.5	32.5	62.6	32.1	45.4	45.3	50.7
Easiness	81.6	64.9	25.8	71.4	29.2	57.8	75.2	68.0	72.7	15.2	46.6	33.8	56.7	57.1	60.9	60.7	24.1	65.4	31.5	43.9	35.3	51.3
CBTS	85.8	65.2	29.4	63.8	31.2	37.2	69.6	64.3	76.2	21.4	56.3	29.8	68.2	60.6	66.2	55.8	30.8	66.1	34.9	48.8	47.1	52.8
WAILS	87.4	86.0	15.5	81.5	35.6	57.9	65.5	60.0	75.0	12.0	67.5	20.4	75.7	60.3	62.8	40.4	35.4	75.1	35.4	60.9	50.9	55.2
AE-PSL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	55.0
MCOF	85.8	74.1	23.6	66.4	36.6	62.0	75.5	68.5	78.2	18.8	64.6	29.6	72.5	61.6	63.1	55.5	37.7	65.8	32.4	68.4	39.9	56.2
DSNA	84.7	73.5	27.1	68.3	43.8	57.2	80.9	64.7	65.4	22.3	69.7	39.2	70.8	73.0	67.8	60.0	45.8	72.6	35.9	57.7	42.2	58.2
DSRG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	61.4
AffinityNet	88.2	68.2	30.6	81.1	49.6	61.0	77.8	66.1	75.1	29.0	66.0	40.2	80.4	62.0	70.4	73.7	42.5	70.7	42.6	68.1	51.6	61.7
Ours1	88.0	76.3	33.7	79.0	48.7	62.5	76.0	69.0	79.5	23.4	66.3	13.7	78.1	69.8	67.3	71.1	39.1	74.4	33.6	51.7	58.8	**60.0**
Ours2	88.1	74.4	30.6	81.4	50.9	69.7	79.9	71.5	86.0	24.0	71.0	18.1	81.0	75.2	65.7	72.2	47.7	78.5	37.4	50.8	61.5	**62.6**

TABLE II

COMPARISON OF DIFFERENT WEAKLY SUPERVISED SEMANTIC SEGMENTATION METHODS ON PASCAL VOC 2012 TEST SET. THE "-" DENOTES UNKNOWN VALUES THAT WERE NOT REPORTED BY THE CORRESPONDING PAPER. THE "OURS1" DENOTES THE OUR NETWORK BASED ON VGG-16 NETWORK WHILE "OURS2" DENOTES THE OUR NETWORK BASED ON RESNET-101 NETWORK

Method	bkg	plane	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	motor	person	plant	sleep	sofa	train	tv	**mIoU**
EM-Adapt	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	39.6
H & S	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	42.6
BFBE	80.3	57.5	24.1	66.9	31.7	43.0	67.5	48.6	56.7	12.6	50.9	42.6	59.4	52.9	65.0	44.8	41.3	51.1	33.7	44.4	33.2	48.0
STC	85.2	62.7	21.1	58.0	31.4	55.0	68.8	63.9	63.7	14.2	57.6	28.3	63.0	59.8	67.6	61.7	42.9	61.0	23.2	52.4	33.1	51.2
SEC	83.5	56.4	28.5	64.1	23.6	46.5	70.6	58.5	71.3	23.2	54.0	28.0	68.1	62.1	70.0	55.0	38.4	58.0	39.9	38.4	48.3	51.7
Easiness	83.0	67.5	29.7	69.7	28.8	59.7	71.2	66.4	69.8	18.6	49.8	44.7	49.4	60.5	73.5	61.8	32.7	62.7	39.0	34.3	36.5	52.8
CBTS	85.7	58.8	30.5	67.6	24.7	44.7	74.8	61.8	73.7	22.9	57.4	27.5	71.3	64.8	72.4	57.3	37.0	60.4	42.8	42.2	50.6	53.7
WAILS	87.7	85.4	14.5	82.5	37.6	58.5	68.6	62.5	73.3	12.3	65.0	20.0	76.1	62.1	63.5	41.0	36.4	77.0	34.9	63.6	51.3	55.9
AE-PSL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	55.7
MCOF	86.8	73.4	26.6	60.6	31.8	56.3	76.0	68.9	79.4	18.8	62.0	36.9	74.5	66.9	74.9	58.1	44.6	68.3	36.2	64.2	44.0	57.6
DSNA	86.3	70.4	29.7	78.8	40.8	53.9	80.6	67.7	67.7	21.0	70.2	46.4	74.8	70.6	74.8	63.7	45.2	76.9	45.0	57.8	39.8	60.1
DSRG	89.3	72.8	32.6	83.3	39.8	68.5	81.1	74.1	81.1	22.4	68.7	23.2	76.8	78.8	79.7	77.4	50.2	80.0	36.5	47.2	63.2	63.2
AffinityNet	89.1	70.6	31.6	77.2	42.2	68.9	79.1	66.5	74.9	29.6	68.7	56.1	82.1	64.8	78.6	73.5	50.8	70.7	47.7	63.9	51.1	63.7
Ours1	88.0	73.9	31.3	75.2	44.9	61.0	77.8	69.6	78.9	24.9	71.9	20.2	74.6	75.0	75.3	74.3	47.1	76.2	45.0	41.1	56.3	**61.1**[1]
Ours2	88.5	79.3	30.2	80.1	43.9	58.4	82.4	77.4	84.8	27.8	74.2	37.4	77.8	76.8	76.6	74.4	55.7	73.4	45.6	41.5	58.3	**64.0**[2]

low-level physical information among superpixels with the feedback chain and generates accurate and concise boundaries. In addition, "baseline + F2" in Table V means not to use the pseudolabels updating mechanism in DFN, which is equal to $w = 0$. When setting $w = 1$, the network fails to predict the semantic segmentation result, and all pixels are predicted as background. It seems that the mechanism updates the pseudolabels too much and the updated pseudolabels become a mess.

Moreover, the "baseline + F1 + F2" indicates integrating both two chains, which rebuild complete position information. The "baseline + F1 + SRW" indicates to use the normal random walk process replacing the standard random walk. It is easy to find that the result is better than "baseline + F1" and worse than "baseline + F2", which indicates the effectiveness of the random walk process and the harmfulness of the unequal confidence problem. The "baseline + F1 + F2-BLF" indicates to remove the boundary loss function in (2). On the one hand, it is obvious to find that the widely used boundary loss term brings a large improvement for weakly supervised semantic segmentation. On the other hand, the second feedback chain achieves a comparable improvement compared with the boundary loss term, demonstrating the effectiveness of our methods. Besides, when we used the second feedback chain, the boundary loss function still works well, indicating the robustness of the proposed method.

Table V shows a 2.4% gain in mIoU when using the first feedback chain. This indicates that our pseudolabels updating mechanism improves the original pseudolabels. To prove the effectiveness of the first feedback chain, we also conduct an experiment that adds this feedback chain to DSRG (based on VGG-16), whose result is 57.8% in mIoU and it has a 0.4% increase compared with our implemented DSRG network. This suggests that our method indeed corrects mislabeled regions hidden in original pseudolabels explicitly. Then, with the comparison of the results between "baseline + F1" and "baseline + F2", we can see that the improvement of the second feedback chain is higher than that of the first feedback chain. It shows that previous works focus on recovering high-level semantic information but pay little attention to rebuilding low-level physical information. By comparing the mIoU value

[1] http://host.robots.ox.ac.uk:8080/anonymous/1ANUBJ.html

[2] http://host.robots.ox.ac.uk:8080/anonymous/2LASXN.html

TABLE III

PERFORMANCE OF DIFFERENT METHODS ON COCO DATASET

Method	BEBP	SEC	WAILS	DSRG	WSIF	Ours
mIoU	20.4	22.4	22.5	26.0	26.3	26.8
f.w. IoU	55.3	60.7	62.7	66.4	67.3	67.6

TABLE IV

COMPARISON OF SEMANTIC SEGMENTATION METHOD UNDER DIFFERENT SUPERVISION TYPES ON PASCAL VOC 2012 SET

Method	Supervision types	val	test
FCN	fully-supervised	-	62.2
DeepLab	fully-supervised	67.6	70.3
WSSL	bounding-box	60.6	62.2
BoxSup	bounding-box	62.0	64.6
RAWK	scribble	61.4	-
ScribbleSup	scribble	63.1	-
What'sPoint	points	46.0	43.6
Ours	image-level tags	62.6	64.0

TABLE V

COMPARISON OF OUR METHOD UNDER DIFFERENT SETTINGS ON PASCAL VOC 2012 VAL SET. "SRW" REPRESENTS THE NORMAL RANDOM WALK PROCESS. "BLF" REPRESENTS THE BOUNDARY LOSS FUNCTION IN (2). THE "ACCU" DENOTES THE PIXEL ACCURACY AND THE "F.W. IOU" DENOTES FREQUENCY WEIGHTED IOU

Method	accu	mIoU	f.w. IoU
Baseline	84.5	52.2	75.3
Baseline + F1	86.6	54.6	76.3
Baseline + F2	88.2	58.4	80.3
Baseline + F1 + SRW	88.40	56.46	79.81
Baseline + F1 + F2	89.3	60.0	81.4
Baseline + F1 + F2 - BLF	85.5	54.5	77.0

of "baseline + F2" with that of MCOF (56.2% based on VGG-16), although both methods attempt to use superpixel, our customized random walk achieves 2.2% performance gain. It suggests that it effectively reduces the over-segmentation phenomenon and reconstructs complete low-level physical information. In addition, we get a larger gain in mIoU with "baseline + F1 + F2" which proves the necessity to rebuild complete position information and the effectiveness of our componentwise approach to reconstructing two categories of position information, respectively.

E. Quality Improvement of Pseudolabels

In this section, we conduct a series of experiments to demonstrate the improvement of pseudolabels.

Table VI reports the mIoU values of generated pseudolabels on PASCAL VOC 2012 val set, against previous methods, namely, SPN [28], DSRG [20] and AffinityNet [29]. The "baseline" represents the original pseudolabels (foreground region from CAM and background region from DRFI), the "superpixel-based" represents the superpixel-based pseudolabels and the "updated" represents the updated pseudolabels. It can be seen that our pseudolabels achieve higher segmentation performance. Although SPN integrates the superpixels into the network structure, our superpixel-based pseudolabels have a 1.7% improvement compared with SPN, which demonstrates the effectiveness to use superpixels as the base unit of pseudolabels. The updating process of DSRG does not correct the initial regions in pseudolabels, and thus, our updated pseudolabels outperform them by 4.3%.

TABLE VI

IMPROVEMENT OF PSEUDOLABELS ON MIOU

Ours			Others		
Baseline	Superpixel-based	Updated	SPN	DSRG	AffinityNet
31.9%	45.5%	61.4%	43.8%	57.1%	59.7%

To explore the relative improvement on pseudolabels using different mechanisms, inspired by the usage of trimap in [56], we also introduce the trimap mask to detect where the majority of improvement takes place. The trimap refers to a narrowband region along object boundaries and the distance of trimap decides the bandwidth, as shown in Fig. 7. In comparison among improvement in trimap of five pixels, improvement in trimap of ten pixels, improvement in trimap of 15 pixels, improvement in trimap of 20 pixels, and improvement of the reversed mask of trimap of 20 pixels, we can clearly see how the two mechanisms work. Meanwhile, using the pseudolabels updating mechanism and the customized superpixel-based random walk mechanism, there exist four different types of pseudolabels. The "original pseudolabels" refers to the generated pseudolabels whose base unit is the square pixels blocks and are the same as DSRG. The "superpixel-based pseudolabels" infers the converted pseudolabels whose base unit is superpixels. The "enhanced pseudolabels" refers to the outputted pseudolabels by our customized random walk process. The "updated pseudolabels" refers to changed pseudolabels by (1).

As shown in Table VII, in comparison between superpixel-based pseudolabels and original pseudolabels, the main improvement occurs in the trimap with a small distance (49.0% when $d \leq 5$ compared with 33.9% when $d > 20$) and demonstrates our superpixel-based pseudolabels indeed produce more accurate segmentation boundaries. In comparison between enhanced pseudolabels and superpixel-based pseudolabels, we can see that the performance improvement is mainly in the regions far away from the boundaries, especially the regions out of trimap of distance less than 20 (10.8% when $d > 20$ compared with 5.3% when $d \leq 20$). It suggests our customized random walk process conquers the over-segmentation phenomenon and focus on removing unnecessary boundaries inner object. In the meantime, the majority of improvement of the updated pseudolabels compared with original pseudolabels takes place in the region inner object, which demonstrates the pseudolabels updating mechanism indeed focuses on the logical structure inner object and recovers high-level semantic information.

F. Effects of Hyperparameters

We finally evaluate the performance of our network with respect to different hyperparameter settings in two feedback chains and specify the way to pinpoint their values.

For the first feedback chain, the hyperparameter w in (3) represents the update rate in the first feedback chain, which

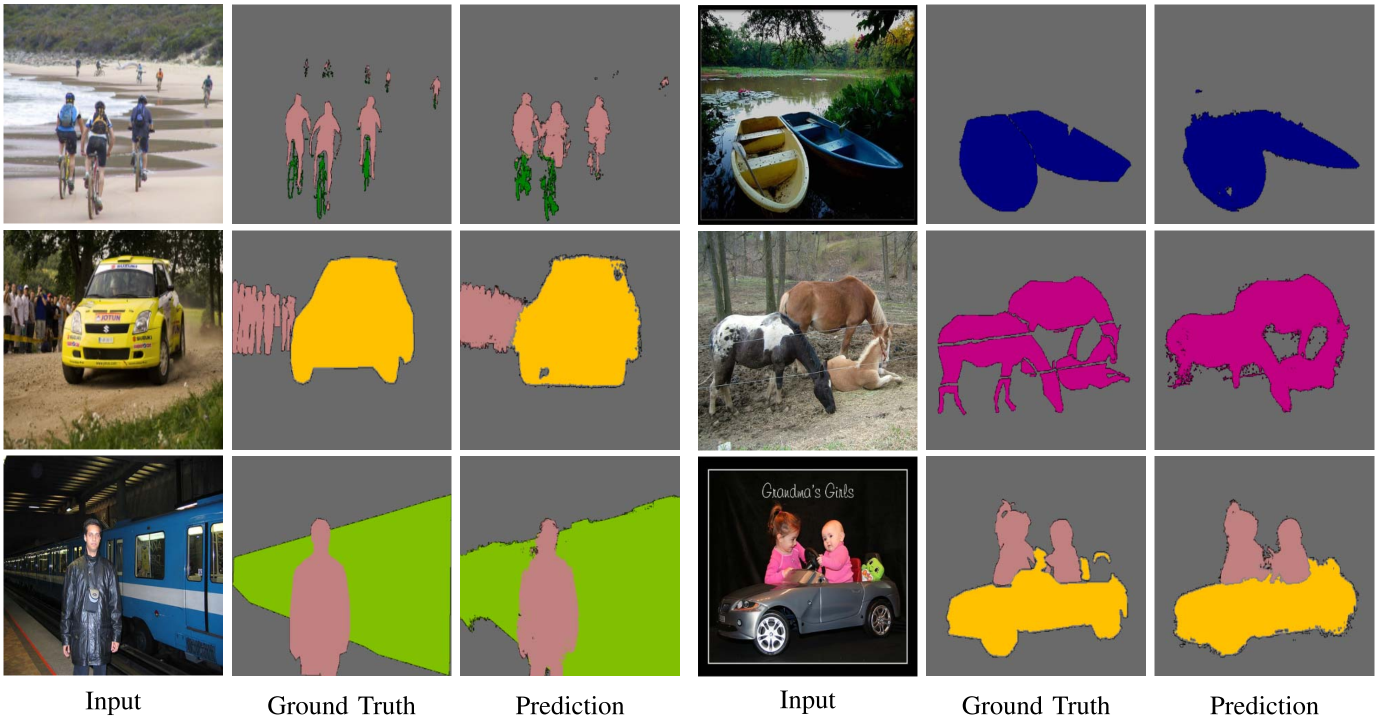


Fig. 6. Qualitative segmentation results on PASCAL VOC 2012 val set. One failure case is shown in the last row.

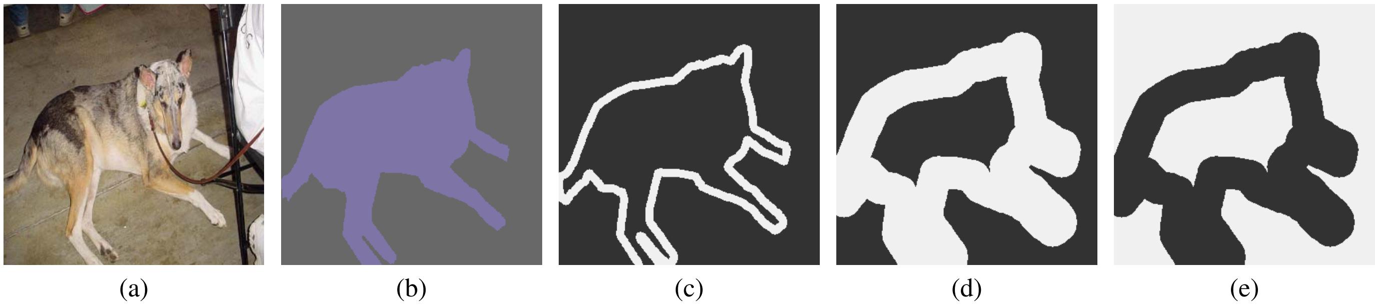


Fig. 7. Some trimap examples. (a) Input image. (b) Ground truth. (c) Trimap of five pixels. (d) Trimap of 20 pixels. (e) Reversed mask of trimap of 20 pixels.

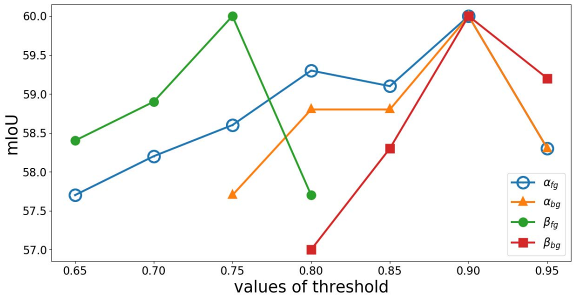


Fig. 8. Performance of our method with respect to different values of α_{fg}, α_{bg}, β_{fg}, and β_{bg}.

determines the weight of feedback from network output. A large w makes pseudolabels change rapidly and cannot supervise network training effectively, whereas a small w slows down the correction of errors incurred by original inaccurate pseudolabels. We go through possible values of w in a large range and select the best-performing one from experiments. From Table VIII, the optimal value of w is 0.20, which reaches the peak point at all three measurement index: accuracy, mIoU, and f.w. IoU.

For the second feedback chain, we also conduct some experiments with different steps counts in a customized random walk process. The steps count n represents the repeating iterations of a single-step customized random walk, which decides the degree of integration of low-level physical information. Results in Table IX suggests that 2 is an optimal point for steps count n. There are no more increases when n is bigger than 2, which shows all low-level physical information is merged into pseudolabels.

In the second feedback chain, there are four hyperparameters, i.e., α_{fg}, α_{bg}, β_{fg}, and β_{bg} in threshold functions. We set their values to $\alpha_{\text{fg}} = 0.90$, $\alpha_{\text{bg}} = 0.90$, $\beta_{\text{fg}} = 0.75$, and $\beta_{\text{bg}} = 0.90$. These hyperparameters are tuned with a coarse-to-fine procedure. In the coarse module, we roughly determine a satisfactory range for each hyperparameter (for example, the range of α_{fg} is $[0.6, 1.0]$). Then, in the fine-tuning module, we divide these parameters into two groups based on their

TABLE VII

RELATIVE IMPROVEMENT OF PSEUDOLABELS ON MIOU. "$d \leq 5$," "$d \leq 10$," "$d \leq 15$," AND "$d \leq 20$" DENOTE EVALUATING USING THE MASK OF TRIMAP OF 5, 10, 15, OR 20 PIXELS. THE "$d > 20$" REPRESENTS USING THE REVERSED MASK OF TRIMAP OF 20 PIXELS. "ALL" DENOTES EVALUATING ON ALL THE PIXELS. THE "S TO O" REPRESENTS THE COMPARISON BETWEEN SUPERPIXEL-BASED PSEUDOLABELS AND ORIGINAL PSEUDOLABELS. THE "E TO S" REPRESENTS THE COMPARISON BETWEEN ENHANCED PSEUDOLABELS TO SUPERPIXEL-BASED PSEUDOLABELS. THE "U TO S" REPRESENTS THE COMPARISON BETWEEN UPDATED PSEUDOLABELS TO SUPERPIXEL-BASED PSEUDOLABELS

Relative Improvement on mIoU	$d \leq 5$	$d \leq 10$	$d \leq 15$	$d \leq 20$	$d > 20$	all
S to O	49.0%	46.9%	40.7%	39.6%	33.9%	42.6%
E to S	4.3%	5.8%	6.4%	5.3%	10.8%	7.9%
U to S	16.5%	20.3%	22.5%	23.0%	48.7%	34.9%

TABLE VIII

PERFORMANCE OF OUR METHOD FOR DIFFERENT w

w	accu	mIoU	f.w. IoU
0.10	88.8	59.1	80.7
0.20	89.3	60.0	81.4
0.30	89.2	59.4	81.4
0.40	89.0	58.7	80.9
0.50	72.3	41.6	60.2
0.60	71.7	41.2	59.8
0.70	76.1	44.2	64.8
0.80	74.9	42.5	63.3
0.90	72.0	41.4	60.1

TABLE IX

PERFORMANCE OF OUR METHOD FOR DIFFERENT n

n	accu	mIoU	f.w. IoU
1	88.9	58.5	80.7
2	89.3	60.0	81.4
3	89.1	59.3	81.2
4	89.0	59.4	80.9

correlations: 1) $\{\alpha_{\text{fg}}, \alpha_{\text{bg}}\}$ and 2) $\{\beta_{\text{fg}}, \beta_{\text{bg}}\}$. When we test values of one group of hyperparameters, another group is set to default values, i.e., the mean value of the optimal range given by the coarse module. In each group, the hyperparameters are tuned with grid search (for example, α_{fg} is tuned at the range [0.6, 1.0] with an interval 0.1). We finally pinpointed the specific value as the one that can achieve the highest mIoU value (for example, the final value of α_{fg} is 0.90). We also test the sensitivity of these hyperparameters. To this end, we evaluate the performance variation for one hyperparameter with the other three fixed. The results are shown in Fig. 8. We can observe that in a wide range, our network outperforms the baseline (52.2%) with a large margin. It demonstrates that our performance is not sensitive to these four hyperparameters. Moreover, the performance variation for α_{fg} and α_{bg} seems less than that for β_{fg} and β_{bg}. One possible reason is that network output changes more rapidly than pseudolabels, which makes the result of thresholding network output more sensitive to its parameters.

V. CONCLUSION

To the best of our knowledge, this article is the first attempt to apply a componentwise approach to recover the lost position information, which is critical to solving the tags supervised semantic segmentation issue. We have explicitly considered the inherent differences between the high-level semantic position information and the low-level physical position information and designed a novel DFN to reconstruct each component independently. Different from the previous methods that aim to recover position information as a whole, we have developed the feedback mechanisms in the tailored network to seek the complete recovery of a separate part of position information. As a result, the generated pseudolabels contain more correctly labeled regions and more accurate boundaries. Extensive experimental results on segmentation datasets have demonstrated the superiority of our approach over the state-of-the-art alternatives in various images.

As for the limitation of this approach, the failure only exists in a few cases, which contain lots of small and dense objects with a complicated background. It is mainly because of the incomplete recovery of low-level physical position information for those images. In the future work, we will explore more on reconstructing the low-level physical position information to address the above-mentioned limitation. Currently, the approach applies a hand-designed relationship matrix to capture the lost low-level physical information, in which some relationships among superpixels are discarded due to the various contexts in natural images. The performance may be further improved if we utilize a deep neural network to generate a more accurate relationship matrix. On the other hand, the proposed approach still uses an encoder-shape network as the backbone, which does not make full use of the details in our pseudolabels that have the same size as input images. Therefore, a further extension of the proposed approach is to apply an encoder–decoder segmentation network, as shown in [57]. Moreover, it would be interesting to introduce generative adversarial nets [58] to recover the missing position information for its amazing ability to reconstruct any data.

REFERENCES

[1] X. Liu, Y.-M. Cheung, M. Li, and H. Liu, "A lip contour extraction method using localized active contour model with automatic parameter selection," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 4332–4335.

[2] L.-T. Law and Y.-M. Cheung, "Color image segmentation using rival penalized controlled competitive learning," in *Proc. Int. Joint Conf. Neural Netw.*, 2003, pp. 108–112.

[3] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[4] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 801–818.

[5] P. Zhang, W. Liu, H. Wang, Y. Lei, and H. Lu, "Deep gated attention networks for large-scale street-level scene segmentation," *Pattern Recognit.*, vol. 88, pp. 702–714, Apr. 2019.

[6] X. Zhu, X. Zhang, X.-Y. Zhang, Z. Xue, and L. Wang, "A novel framework for semantic segmentation with generative adversarial network," *J. Vis. Commun. Image Represent.*, vol. 58, pp. 532–543, Jan. 2019.

[7] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Feb. 17, 2021, doi: 10.1109/TPAMI.2021.3059968.

[8] J. Shen, X. Dong, X. Jin, L. Shao, F. Porikli, and J. Peng, "Submodular function optimization for motion clustering and image segmentation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2637–2649, Sep. 2019.

[9] Q. Peng *et al.*, "Regularized-ncut: Robust and homogeneous functional parcellation of neonate and adult brain networks," *Artif. Intell. Med.*, vol. 106, Jun. 2020, Art. no. 101872.

[10] J. Shen, J. Peng, X. Dong, L. Shao, and F. Porikli, "Higher order energies for image segmentation," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4911–4922, Oct. 2017.

[11] X. Dong, J. Shen, L. Shao, and M.-H. Yang, "Interactive cosegmentation using global and local energy optimization," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3966–3977, Nov. 2015.

[12] J. Dai, K. He, and J. Sun, "BoxSup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1635–1643.

[13] P. Tang, X. Wang, Z. Huang, X. Bai, and W. Liu, "Deep patch learning for weakly supervised object classification and discovery," *Pattern Recognit.*, vol. 71, pp. 446–459, Nov. 2017.

[14] Y. Zhang, Y. Bai, M. Ding, Y. Li, and B. Ghanem, "Weakly-supervised object detection via mining pseudo ground truth bounding-boxes," *Pattern Recognit.*, vol. 84, pp. 68–81, Dec. 2018.

[15] D. Lin, J. Dai, J. Jia, K. He, and J. Sun, "ScribbleSup: Scribble-supervised convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 3159–3167.

[16] W. Wu, H. Qi, Z. Rong, L. Liu, and H. Su, "Scribble-supervised segmentation of aerial building footprints using adversarial learning," *IEEE Access*, vol. 6, pp. 58898–58911, 2018.

[17] A. Bearman, O. Russakovsky, V. Ferrari, and L. Fei-Fei, "What's the point: Semantic segmentation with point supervision," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 549–565.

[18] G. Papandreou, L.-C. Chen, K. P. Murphy, and A. L. Yuille, "Weakly- and semi-supervised learning of a deep convolutional network for semantic image segmentation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1742–1750.

[19] A. Kolesnikov and C. H. Lampert, "Seed, expand and constrain: Three principles for weakly-supervised image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 695–711.

[20] Z. Huang, X. Wang, J. Wang, W. Liu, and J. Wang, "Weakly-supervised semantic segmentation network with deep seeded region growing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7014–7023.

[21] X. Liu, Z. Hu, H. Ling, and Y.-M. Cheung, "MTFH: A matrix tri-factorization hashing framework for efficient cross-modal retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 3, pp. 964–981, Mar. 2021.

[22] X. Wang, S. You, X. Li, and H. Ma, "Weakly-supervised semantic segmentation by iteratively mining common object features," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1354–1362.

[23] F. S. Saleh, M. S. Aliakbarian, M. Salzmann, L. Petersson, J. M. Alvarez, and S. Gould, "Incorporating network built-in priors in weakly-supervised semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1382–1396, Jan. 2018.

[24] Y. Wei *et al.*, "STC: A simple to complex framework for weakly-supervised semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 11, pp. 2314–2320, Nov. 2017.

[25] T. Zhang, G. Lin, J. Cai, T. Shen, C. Shen, and A. C. Kot, "Decoupled spatial neural attention for weakly supervised semantic segmentation," *IEEE Trans. Multimedia*, vol. 21, no. 11, pp. 2930–2941, Nov. 2019.

[26] H. Zhou, K. Song, X. Zhang, W. Gui, and Q. Qian, "WAILS: Watershed algorithm with image-level supervision for weakly supervised semantic segmentation," *IEEE Access*, vol. 7, pp. 42745–42756, 2019.

[27] W. Shimoda and K. Yanai, "Weakly supervised semantic segmentation using distinct class specific saliency maps," *Comput. Vis. Image Understand.*, vol. 191, Feb. 2020, Art. no. 102712.

[28] S. Kwak, S. Hong, and B. Han, "Weakly supervised semantic segmentation using superpixel pooling network," in *Proc. AAAI Conf. Artif. Intell.*, 2017, vol. 31, no. 1, pp. 4111–4117.

[29] J. Ahn and S. Kwak, "Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 4981–4990.

[30] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2017.

[31] D. Pathak, E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional multi-class multiple instance learning," in *Proc. 3rd Int. Conf. Learn. Represent.*, 2015, pp. 1–4.

[32] D. Pathak, P. Krahenbuhl, and T. Darrell, "Constrained convolutional neural networks for weakly supervised segmentation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1796–1804.

[33] Y. Wei, J. Feng, X. Liang, M. M. Cheng, Y. Zhao, and S. Yan, "Object region mining with adversarial erasing: A simple classification to semantic segmentation approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1568–1576.

[34] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2921–2929.

[35] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," in *Proc. 2nd Int. Conf. Learn. Represent.*, 2014, pp. 1–8.

[36] W. Shimoda and K. Yanai, "Weakly supervised semantic segmentation using distinct class specific saliency maps," *Comput. Vis. Image Understand.*, vol. 191, Feb. 2020, Art. no. 102712.

[37] J. Wang, H. Jiang, Z. Yuan, M.-M. Cheng, X. Hu, and N. Zheng, "Salient object detection: A discriminative regional feature integration approach," *Int. J. Comput. Vis.*, vol. 123, no. 2, pp. 251–268, 2017.

[38] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, Nov. 2006.

[39] X. P. Dong, J. B. Shen, L. Shao, and L. van Gool, "Sub-Markov random walk for image segmentation," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 516–527, Feb. 2016.

[40] J. Shen, Y. Du, W. Wang, and X. Li, "Lazy random walks for superpixel segmentation," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1451–1462, Apr. 2014.

[41] F. Liu, G. Lin, R. Qiao, and C. Shen, "Structured learning of tree potentials in CRF for image segmentation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2631–2637, Jun. 2018, doi: 10.1109/TNNLS.2017.2690453.

[42] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[43] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, Sep. 2004.

[44] R. Hettiarachchi and J. F. Peters, "Voronoï region-based adaptive unsupervised color image segmentation," *Pattern Recognit.*, vol. 65, pp. 119–135, May 2017.

[45] M. Mostajabi, P. Yadollahpour, and G. Shakhnarovich, "Feedforward semantic segmentation with zoom-out features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3376–3385.

[46] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent.*, 2015, pp. 1–14.

[47] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes challenge: A retrospective," *Int. J. Comput. Vis.*, vol. 111, no. 1, pp. 98–136, Jan. 2014.

[48] T. Y. Lin *et al.*, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.

[49] B. Hariharan, P. Arbeláez, L. Bourdev, S. Maji, and J. Malik, "Semantic contours from inverse detectors," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 991–998.

[50] A. Roy and S. Todorovic, "Combining bottom-up, top-down, and smoothness cues for weakly supervised image segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3529–3538.

[51] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[52] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected CRFs with Gaussian edge potentials," in *Advances in Neural Information Processing Systems*. Granada, Spain: Curran Associates, vol. 24, 2011, pp. 109–117.

[53] C. Redondo-Cabrera, M. Baptista-Rios, and R. J. Lopez-Sastre, "Learning to exploit the prior network knowledge for weakly supervised semantic segmentation," *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3649–3661, Jul. 2019.

[54] Y. Li, Y. Liu, G. Liu, and M. Guo, "Weakly supervised semantic segmentation by iterative superpixel-CRF refinement with initial clues guiding," *Neurocomputing*, vol. 391, pp. 25–41, May 2020.

[55] P. Vernaza and M. Chandraker, "Learning random-walk label propagation for weakly-supervised semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7158–7166.

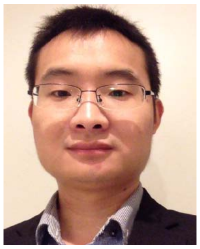
[56] T. Wang, Z. Ji, Q. Sun, Q. Chen, Q. Ge, and J. Yang, "Diffusive likelihood for interactive image segmentation," *Pattern Recognit.*, vol. 79, pp. 440–451, Jul. 2018.

[57] J. Liu, Y. Wang, Y. Li, J. Fu, J. Li, and H. Lu, "Collaborative deconvolutional neural networks for joint depth estimation and semantic segmentation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5655–5666, Nov. 2018, doi: 10.1109/TNNLS.2017.2787781.

[58] I. J. Goodfellow *et al.*, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, vol. 27. Montreal, QC, Canada: MIT Press, 2014, pp. 2672–2680.

Zhengqiang Zhang received the B.S. and M.S. degrees with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China, in 2017 and 2020, respectively.

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



Qinmu Peng received the Ph.D. degree from the Department of Computer Science, Hong Kong Baptist University, Hong Kong, in 2015.

He is currently an Assistant Professor with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, Hubei, China. His current research interests include medical image processing, pattern recognition, machine learning, and computer vision.

Sichao Fu is currently pursuing the Ph.D. degree with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China.

His research interests include pattern recognition and deep manifold learning.

Wenjie Wang received the B.E. degree with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China, in 2019, where he is currently pursuing the M.Sc. degree.

His current research interests include multimodal retrieval, zero-shot learning, and causal learning.

Yiu-Ming Cheung (Fellow, IEEE) is currently a Full Professor with the Department of Computer Science, Hong Kong Baptist University, Hong Kong. His current research interests include machine learning, pattern recognition, visual computing, and optimization.

Dr. Cheung is a fellow of the Institution of Engineering and Technology (IET), the British Computer Society (BCS), the Royal Society of Arts (RSA), and a Distinguished Fellow of IETI. He is the Founding Chair of the Computational Intelligence Chapter of the IEEE Hong Kong Section. He serves as an Associate Editor for the IEEE TNNLS, *Cybernetics*, *Pattern Recognition* (PR), and *Knowledge and Information Systems*.

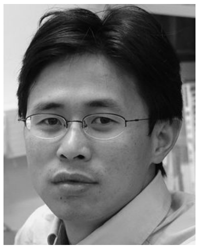
Yue Zhao received the Ph.D. degree from the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China, in 2019.

She is currently an Assistant Professor with the School of Computer Science and Information Engineering, Hubei University, Wuhan. Her research interests include computer vision, pattern recognition, and machine learning.

Shujian Yu (Member, IEEE) received the B.S. degree from the School of Electronic Information and Communications, Wuhan, China, in 2013, and the Ph.D. degree in electrical and computer engineering with a minor in statistics from the University of Florida, Gainesville, FL, USA, in 2019.

He is currently an Associate Professor with the Machine Learning Group, UiT The Arctic University of Norway, Tromsø, Norway. His research interests include machine learning, information theory, and signal processing.

Dr. Yu was a recipient of the 2020 International Neural Networks Society's Aharon Katzir Young Investigator Award.



Xinge You (Senior Member, IEEE) received the B.S. and M.S. degrees in mathematics from Hubei University, Wuhan, China, in 1990 and 2000, respectively, and the Ph.D. degree from the Department of computer Science, Hong Kong Baptist University, Hong Kong, in 2004.

He is currently a Professor with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan. His current research interests include image processing, wavelet analysis and its applications, pattern recognition, machine earning, and computer vision.