Generic Image Segmentation in Fully Convolutional Networks by Superpixel Merging Map

Anonymous ACCV 2020 submission

Paper ID 938

Abstract. Recently, the Fully Convolutional Network (FCN) has been widely adopted in image segmentation. However, existing FCN-based segmentation algorithms were designed for semantic segmentation. It is hard to integrate the FCN into generic image segmentation. Before learning-based segmentation algorithms were developed, many advanced generic segmentation algorithms are superpixel-based. Unfortunately, due to the irregular shape and size of superpixels, there is no generic image segmentation algorithm that utilize both superpixels and the FCN. In this paper, we combined the merits of the FCN and superpixels and proposed a highly accurate and extremely fast generic image segmentation algorithm. We treated image segmentation as multiple superpixel merging decision problems and determined whether the boundary between two adjacent superpixels should be kept. In other words, if the boundary of two adjacent superpixels should be deleted, then the two superpixels will be merged. The proposed algorithm takes a multichannel image consisting of stacked color image and several feature maps such as the superpixel boundary map and the edge detection result as the input of a deep neural network. The network outputs the predicted superpixel boundary map that indicates whether the boundary of two adjacent superpixel should be keep, in a way, making decision about merging suprepixels. That is, by solving all the superpixel-merging subproblems with just one forward pass, the FCN facilitates the speed of the whole segmentation process by a wide margin meanwhile gaining higher accuracy. Additionally, with the use of multiple superpixel generation algorithms, one can obtain a huge amount of training data by varying the parameters of different superpixel algorithms. That is, the proposed method has low requirement for human-annotated groundtruth. Simulations show that the proposed algorithm has favorable runtime, meanwhile achieving highly accurate segmentation results. It outperforms state-of-the-art image segmentation methods, including feature-based and learning-based methods, in all evaluation metrics.

1 Introduction

Image segmentation is fundamental and important in many image processing applications. There are many existing image segmentation algorithms, including the region growing method [1], the mean shift method [2], the watershed [3,4], the normalized cut [5,6], the graph-based method [7], and the
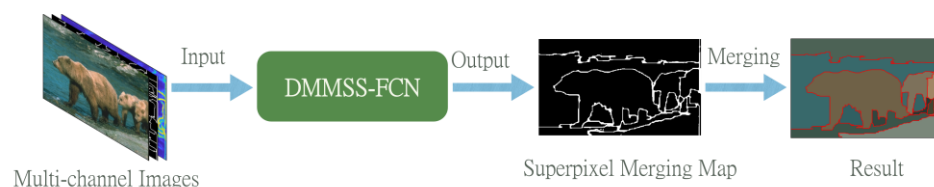
2   ACCV-20 submission ID 938

Fig. 1. Overview of the proposed *DMMSS-FCN* algorithm. Multiple feature maps including an original RGB image are stacked together to form the input data. Then, they are fed into the fully convolutional network. The output is the superpixel merging map that determines which boundaries should be kept. With such information, we perform superpixel merging based on the output of the model to produce the final segmentation result.

superpixel-based method [8–10].

In recent years, deep learning techniques have also been adopted in image segmentation [11–14]. With supplicated deep learning architectures, one can achieve good segmentation results with enough training time.

However, these learning-based image segmentation algorithms are used to produce semantic segmentation but not generic segmentation result. Before learning-based segmentation algorithms were developed, many advanced image segmentation algorithms are based on superpixels. However, due to the irregularity of sizes and shapes of superpixels, it is hard to apply superpixels in a learning-based generic segmentation architecture. Therefore, we broke down the whole image segmentation problem into several superpixel merging decision problems. Furthermore, each superpixel merging process can be converted into a boundary keeping problem. That is, by keeping the boundary between two adjacent superpixels, in a way, we have already decided those superpixels should be merged or not.

In this study, we proposed a novel algorithm that leverages the Fully Convolutional Network (FCN) for learning generic image segmentation. We call it Deep Merging Models for Superpixel-Segmentation by Fully Convolutional Networks (DMMSS-FCN). With the use of FCN, all of those superpixel merging problems can be solved in just one forward pass. That is, it is extremely efficient. First, the proposed *DMMSS-FCN* model will predict whether the pixels along the boundary of two adjacent superpixels should be keep. Following, a majority voting technique will be applied to decide the existence of all superpixel boundaries. Therefore, the final image segmentation result will be produced with minimum effort.

We use 5-channel stacked images as the input, including an RGB image, a superpixel boundary map, an edge detection map. Since different superpixel algorithms will produce different superpixel boundary maps, by vary the parameters of different superpixel algorithms, numerous training data can be easily obtained. That is, our method does not require many human-annotated ground truth. Furthermore, our method is fully automatic, that is, user does not need to assign number of regions in prior.

We show the overview of the proposed *DMMSS-FCN* algorithm in Fig. 1. We will discuss the detail of *DMMSS-FCN* in Chapter 3.

2 Related Work

2.1 Superpixels

Superpixels are a group of pixels with similar colors and locations. There are many types of superpixels, including the entropy rate superpixel (ERS) [15] and the simple linear iterative clustering (SLIC) [16] superpixel. In [2], the superpixel generated by mean shift was proposed. It has good edge-preserving property and the number of superpixels have not to be specified in advance. Moreover, its boundaries highly match the borders of objects. Recently, deep learning-based superpixels like superpixel sampling network (SSN) [17], and segmentation-aware loss (SEAL) superpixels [18] were proposed. They both outperform non deep learning-based superpixels by a wide margin.

2.2 Classical Segmentation Algorithms

Most classical segmentation methods utilize hand-crafted features such as color, histogram, gradient, or texture to perform segmentation. Many of them are still widely used today, such as the graph-based method [7] and the normalized cut [6]. Arbelaz *et. al.* [3,4] proposed a method based on the global information to perform the oriented watershed transform and generated an ultra-metric contour map for hierarchical segmentation.

Moreover, superpixel-based segmentation algorithms like the method of segmentation by aggregating superpixels (SAS) [8] perform segmentation based on merging superpixels using some local grouping cues. Kim *et. al.* [9] used a full range affinity model and Yang *et. al.* [10] proposed a spectral clustering method based on Gaussian Kernel similarity measure for image segmentation. These superpixel-based segmentation algorithms have good performance. However, due to the irregular shapes and sizes of superpixels, it is hard to embed deep learning techniques in superpixels-based algorithms.

2.3 Deep Learning in Image Segmentation

Recently, many semantic segmentation algorithms applied the deep neural network were developed. In [12,13], the fully convolutional network (FCN) were proposed to improve the performance of image segmentation and object detection. In [14], the conditional random field (CRF) was applied in the pixel-wise segmentation method of DeepLab. In [11], Xia and Kulis proposed the W-Net based on the FCN to perform segmentation. In [19], Haeh *et. al.* introduced a method based on the CNN to detect split error in segmented biomedical images.

In this study, we integrate the merits of superpixel-based methods and state-of-the-art learning-based methods and propose a high accuracy image segmentation algorithm.

4   ACCV-20 submission ID 938

3 Proposed Algorithms: DMMSS-FCN

In this section, we proposed an effective way of integrating FCN with generic image segmentation. That is, with the use of superpixel, we can encode the superpixel merging problems with keeping of superpixel boundaries into generic image segmentation. In Fig. 2a, the existence of boundary between two superpixels implies that they are separated from each other, in other words, they are not merged. On the contrary, the disappearance of boundary means they are merged. Therefore, with the use of the FCN, we perform dense prediction on every adjacent boundaries. After one forward pass in the FCN, all the pixels along the boundaries are classified into two classes, *keeping* or *removing*. Then we measure the tendency of keeping a boundary by majority voting technique defined in (1) for quantization. Hence, we can easily recover the segmentation result.

$$BoundaryRate(i, j) = \frac{\# \ of \ pixels \ of \ \left(keeping \ label \ \cap Bnd(i, j) \right)}{\# \ of \ pixels \ of \ Bnd(i, j)} \tag{1}$$

where $keep$ label indexes the label of pixel prediction on the boundary pixel is to keep the boundary. For example in Fig. 2b, there are 15 pixels along the boundary $Bnd(i, j)$, 10 of them are predicted as $keeping$ labels (orange circles) while 5 of them (black circles) are $removing$ labels, resulting in a $BoundaryRate$ of 2/3. Therefore, we can thresholding on the $BoundaryRate$ to get the segmentation result from the FCN output.

3.1 Five-channel Input Data

We defined our input data by concatenating a RGB (3 channels) image with a superpixel boundary map (1 channel) and a edge-detection map (1 channel). The RGB image indicates the original image while the superpixel boundary map is a binary image with only the boundary between two adjacent superpixels are marked as positive (the $keeping$ label). The superpixel boundary map generation is shown in Fig. 3. The RefineContourNet (RCN) [20] is adopted for edge detection. An example of the output of the RCN is shown in the right subfigure of Fig. 3.

We have also tried different combination of concatenating input images, ablation studies have been carried out to analyze the impact of different input channel in Sec. 4.3.

Since we want our model to be adaptive to any input superpixel type, the superpixel boundary map plays an important role in the whole process. As a result, it is essential to our input data. That is, with any superpixel result as prior input, our model can determine which boundary should be keep, with the superpixel boundary map as attention mechanism.
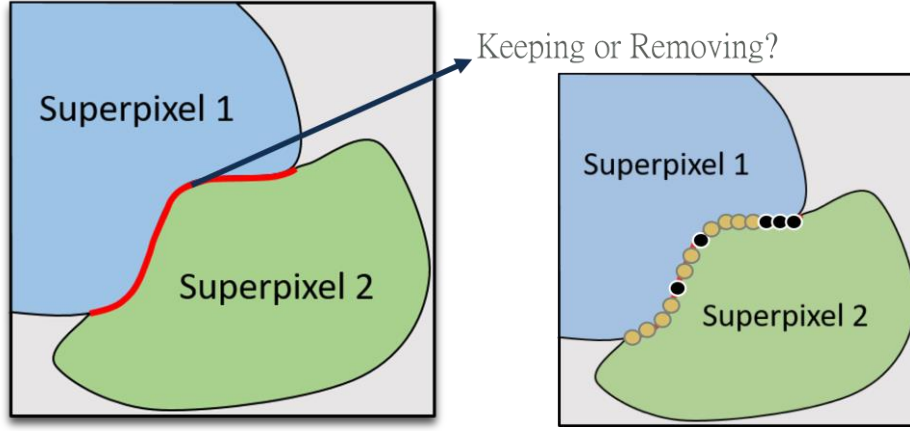
Fig.2. Superpixel merging transformation. **Left:** Converting superpixel merging problem into superpixel boundary keeping problem. **Right:** An example of the boundary rate.
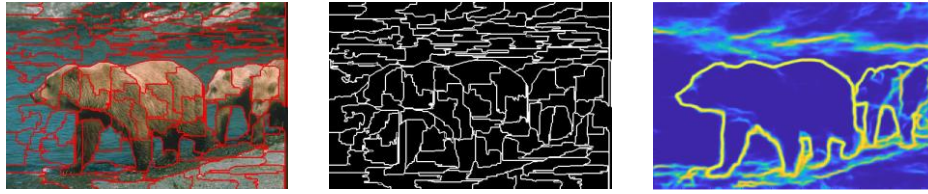

Fig.3. Component of input data. **Left:** Superpixel result from SEAL-ERS. **Middle:** Binary image with superpixel boundaries marked as 1s, others are 0s. **Right:** Edges extracted by the RCN.

3.2 Ground Truth Generation and Output

In [3], the Segmentation Covering (SC), the Probabilistic Rand Index (PRI), and the Variation of Information (VI) are proposed to be the standard evaluation metrics for generic image segmentation. Among all, the PRI in (2) are often treated as the most important metric. That is, we use an oracleguided process to produce the highest achievable PRI score for a given oversegmentation(superpixel) result. From that, we can acquire the groundtruh by transforming the results into a binary boundary map called the *superpixel merging map* which is the ideal output of our model. The groundtruth generation result is shown in Fig. 4 Since different superpixel algorithms or parameters could yield different groundtruth pairs. Sufficient training data can be obtained by varying the parameters of superpixel generation algorithms. Therefore, the model is designed to be adaptive for different superpixel type. The model should response to different superpixel boundary map as prior input. That is, we want the model to output superpixel merging map that adheres to the input superpixel boundaries, since we use *BoundaryRate* to recover segmentation result, bad localization of dense prediction could cripple the effectiveness of *BoundaryRate*. In Fig. 5, we show that for different superpixel type as prior input, our models will generate corresponding superpixel merging maps.

6   ACCV-20 submission ID 938

Fig. 4. Ground truth generation. **Left:** SEAL-ERS superpixel result. **Middle:** Perfect segmentation by oracle. **Right:** Ground-truth superpixel boundary map by converting the resulting ground truth into a binary image.
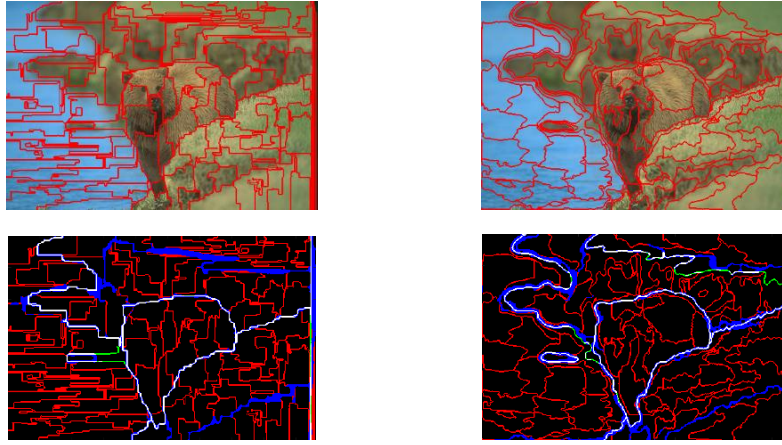

Fig. 5. Different superpixel boundary map input. (1st **Row):** Superpixel boundaries overlapped with the original images. (2nd **Row):** Superpixel merging maps overlapped with corresponding ground-truth. The blue lines are the preserved boundaries, the white lines are the boundary in the ground truth, and the red lines are the removed boundaries.

$$PRI\left(S,\{G_k\}\right)=\frac{1}{T}\sum_{i<j}\left[c_{ij}p_{ij}+\left(1-c_{ij}\right)\left(1-p_{ij}\right)\right] \qquad (2)$$

where c_{ij} indicates the case where pixels i and j belong to the same region, and p_{ij} is the probability. S is the resulting segmentation, and G_k is a set of ground-truths.

3.3 Training Architecture

Since our goal is to perform pixel-wise prediction on the image, and examine the *BoundaryRate* along the boundary of two adjacent superpixels, the localization of predicted labels is crucial in our task. Therefore, we adopted the DeepLab V3+[21] FCN architecture proposed by Chen *et. al.*, it utilized traditional Encoder-Decoder architecture for semantic segmentation with atrous convolution implemented in the encoder side for better field of view, and an simple but effective decoder with short cut skipped through the encoder part is added to this architecture, making it a highly accurate Fully Convolutional Networks structure while preserving good spacial information.

ACCV2020
#938

ACCV2020
#938

CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

ACCV-20 submission ID 938   7

In this paper, we adopted the InceptionResNetV2[22] as the hidden encoder architecture, and the output stride of encoder is set to 16. And the batch size is set to 13 with the size of input data is 321×481×5.

Since this is a binary classification problem of two labels, *keeping* or *removing* labels, with the *keeping* class is far less than the other one, we applied the weighted binary cross entropy loss in (3) to calculate the loss. Furthermore, Adam optimizer is adopted to update the parameters of the networks with initial learning rate of 0.0001 and divided by 0.5 for every 10 epochs, and stopped training after 100 epochs.

$$weighted - BCE(T,S) = -\sum_i \left[w_0 T_i \log(S_i) + w_1 (1-T_i) \log(1-S_i) \right] \qquad (3)$$

where S is the output superpixel merging map, and $S_i \in (0,1)$ denotes the predicted probability value at i-th pixel in S. T is the groundtruth merging map, and $T_i \in \{0,1\}$ denotes the groundtruth label in T. The weighting factors w_0 and w_1 is the ratio of two classes.

We applied the Berkeley Segmentation DataSet 500, with 200 test images, 200 training images, and 100 validation images. As we mentioned in Sec 3.2, different superpixel algorithm can produce different superpixel and groundtruth pairs. Therefore, we adopted SEAL-ERS [18] and SSN [17] both with the number of superpixels set to 100, and 200 to generate the training data, resulting in 800 training images, and 400 validation images in total.

3.4 Inference and Superpixel Merging

As we shown in Fig. 6, we first concatenate RGB image with superpixel boundary map and RCN edge map to form the 5-channel input data. Then we perform one forward pass through the networks to obtain superpixel merging map. Afterwards, we use the *BoundaryRate* in (1) to measure how many predicted *keepings* labels are on each boundary of adjacent superpixels. Then, we perform thresholding merging procedure as we by using the formula in (1). Here we adopted an adaptive thresholding technique which first start merging by thresholding with the lowest threshold values and increase the threshold values bit-by-bit until there are no candidates for merging. After the thresholding value is reached to our pre-defined value (0.5 in this paper), the whole merging process stops and output the final segmentation result.

4 Experiments

In this section, we carried out tons of experiments and ablation studies to justify that our proposed *DMMSS-FCN* can outperform many state-of-the-art algorithms, even compared to our own *DMMSS*, the proposed *DMMSS-FCN* can still surpass it in all evaluation metrics while boosting the speed, making *DMMSSFCN* a highly accurate and efficient generic image segmentation algorithm.

4.1 Segmentation Evaluation

8   ACCV-20 submission ID 938

Table 1. Results on the BSDS500 dataset.

Method	VI	PRI	SC
Ncuts	2.23	0.78	0.45
Canny-owt-ucm	2.19	0.79	0.49
Felz-Hutt	2.21	0.80	0.52
Mean Shift	1.85	0.79	0.54
Taylor	1.78	0.81	0.56
W-Net	1.76	0.81	0.57
fPb-owt-ucm	1.70	0.82	0.58
DC-Seg-full	1.68	0.82	0.59
W-Net+ucm	1.67	0.82	0.59
gPb-owt-ucm	1.69	0.83	0.59
cPb-owt-ucm	1.65	0.83	0.59
DMMSS-FCN(SSN)	**1.38**	**0.87**	**0.66**
Human Drawing	1.17	0.88	0.72

Table 2. Runtime result on the BSDS500 test set.

Method	Mode	Time (s)
DMMSS-FCN(SEAL-ERS)	CPU+GPU	**0.26**
DMMSS-FCN(SEAL-ERS)	CPU	**8.2**
DC-Seg-full	CPU	59
gPb-OWT-UCM	CPU	100

To compare the proposed *DMMSS-FCN* algorithm to the existing methods, we evaluate the performance on the standard metrics of segmentation covering (SC), the probabilistic rand index (PRI), and the variation of information (VI) [3] same as we adopted in the previous chapter. A higher SC and PRI and a lower VI mean better performance.

We compare the proposed *DMMSS-FCN* algorithm to the state-of-the-art methods, including the W-Net [11], gPb-owt-ucm [4], DC-Seg-full [23], Taylor [24], Felzenszwalb and Huttenlocher (Felz-Hutt) [7], Mean Shift [2], Canny-owtucm [4], Multiscale Normalized Cuts (NCuts) [5], fPb-owt-ucm [9], cPb-owtucm [9], and our own *DMMSS*. As the proposed algorithm, all the algorithms compared in Table 1 have not to assign the number of regions in prior.

4.2 Run Time Analysis

We then analyze the run time of proposed the *DMMSS-FCN* and compare it to the state-of-the-art segmentation algorithms of DC-Seg-full [23] and gPb-OWT-UCM [4]. We evaluate our algorithms on SEAL-ERS superpixel, and perform inference on the BSDS500 test set.

ACCV2020
#938

ACCV2020
#938

CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

ACCV-20 submission ID 938   9

Table 3. Performance of different input data combination.

Input Data	VI	PRI	SC
4-channel: RGB(3)+spixel bdry(1)	1.569	0.846	0.547
5-channel: RGB(3)+spixel bdry(1) +RCN edge(1)	**1.446**	**0.863**	**0.634**
5-channel: LAB(3)+spixel bdry(1) +RCN edge(1)	1.458	0.860	0.632
7-channel: RGB(3)+spixel bdry(1) +RCN edge(1)+Affinity XY(2)	1.472	0.861	0.629
7-channel: LAB(3)+spixel bdry(1) +RCN edge(1)+Affinity XY(2)	1.534	0.859	0.606

In Table 2, the average run time of processing single image is presented. One can find out that the proposed *DMMSS-FCN* has drastically reduce the run time compare to gPb-OWT-UCM, decreasing the runtime up to 384x times less than the gPb-OWT-UCM and 227x times less than DC-Seg-full and meanwhile achieving higher accuracy. More on that, even we switch our *DMMSS-FCN* to CPU mode, our proposed *DMMSSFCN* still 12x faster than the gPb-OWT-UCM and 12x faster than the DC-Segfull.

4.3 Ablation Study

In this section, we will discuss the functionality of some key component in our proposed algorithm, and how they affect the overall performance. Including the different combination of input data, different superpixel generation algorithm, and different DeepLabV3 Plus implementation detail, furthermore, different inference techniques.

Combination of Input Data

Following, we show that different combination of input data could have a great impact on the performance of the model. In this section, we mainly use the SEAL-ERS 100 as the underlying superpixel representation. In Table 3, we show the difference of concatenating different feature map as input data.

In the early design stage of our proposed *DMMSS-FCN*, we first design the input data to have only 4-channel, that is, a RGB image(3-channel) concatenated with superpixel boundary map. During training, we found out that the model did not perform as we expected. As a result, we thought that concatenating another feature map as prior information might improve the performance. Therefore, as we added the RCN edge-detection map, we got a huge gain in performance. Hence, we tried to concatenated more feature map or change RGB to CIE LAB color space to further improve the performance. For example, we use the Affinity map that generated from the SEAL-ERS as candidates for feature map concatenation. Nevertheless, from the experimental results in Table 3, more channels of feature maps do not imply better performance. In the end, we adopted the 5-channel with RGB image as the input of the model.
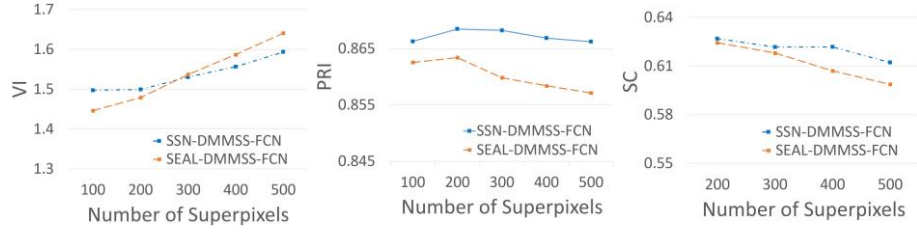
10   ACCV-20 submission ID 938

Fig. 7. Performance of different numbers of superpixels between SSN and SEALERS over *DMMSS-FCN*.

Superpixels

We measure the performance over different number of initial superpixels. In Fig. 7, the proposed *DMMSS-FCN*. Since our main superpixel boundary maps are collected from the initial number of superpixels set to 100, and 200, the peak performance is around 200. However, our proposed *DMMSS-FCN* still maintains great performance over different numbers of initial superpixels even if some of them are not included during training.

In this work, we input the superpixel boundary map as attention mechanism for model to focus on the boundary part then make prediction on the pixels along boundary. In Fig. 5, we show the performance of the proposed model on different types of superpixels, where yellow lines index that the prediction overlapped with the corresponding ground-truth superpixel boundary map, the light blue parts indicate that the model predict *keeping* label, and the dark blue parts represent the model predict *removing* label. Therefore, as we observe, different input superpixel type will lead the model to predict corresponding superpixel merging map.

DeepLabV3 Plus Variation

In [21], Chen *et. al.* proposed a powerful encoder-decoder-based fully convolutional networks. They suggested different variations of implementation details, including changing the output stride of the model or replacing hidden encoder architecture with any existing CNN architecture such as the ResNet [25] or Xception [26]. More on that, inference techniques like forwarding not one image but four up-down and right-left flipped images at once is also reported as legit method to gain performance. Therefore, the following article will discuss the difference of the performance among the variation of DeepLabV3 Plus applied in our work.

We choose SEAL-ERS 100 as the underlying superpixel representation to carry out the following experiments.

In Table 4, we reported the difference among two recent CNN architecture, the Xception [26] and the InceptionResNetV2 [22] with two different output stride setting. In the original DeepLabV3 paper [14], the author indicated that smaller output stride could greatly increase the training time and memory usage with a negligible improvement which corresponds to our case here. As for different encoder architecture, InceptionResNetV2 has better performance over the Xception.

ACCV2020
#938

ACCV2020
#938

CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

ACCV-20 submission ID 938   11

Table 4. Performance of different encoder architecture and output stride.

Encoder Architecture	Output Strides	VI	PRI	SC
Xception	8	1.569	0.860	0.600
Xception	16	1.532	0.857	0.606
InceptionResNetV2	8	1.635	0.854	0.583
InceptionResNetV2	16	1.446	0.863	0.634

Furthermore, we also adopted the inference technique in [21] to improve the performance, that is, we generate extra images by flipped the original image, then feed them into the networks, then average the scores of the outputs to obtain the final superpixel merging map. In Table 5, we show that a simple but effective flipping technique could boost the performance.

4.4 Visual Comparison

In this section, visual comparisons with other segmentation algorithms are presented. In addition to show the simulation results on the BSDS500, we also perform simulations on some real-world images to justify that under any circumstance, our proposed algorithm can be highly accurate and efficient to produce good generic image segmentation results.

BSDS500 Test Images

Fig. 9 shows the comparison between the proposed methods and other methods like a deep-learning-based method, the W-Net [11], and classical segmentation algorithm gPb-OWT-UCM [4]. As we can see, both the results of the proposed algorithms are much better than that of state-of-the-art algorithms, since ours can produce more general and compact segmentation results compared to the others.

Real-World Images

To further justify the robustness of our model, we pick some modern images that have never been in our training set. In this section, we show some segmentation results on the images taken in the night to test the capability of our proposed methods of handling dark view scenario. In Fig. 8a ~ 8b, one can see that, the proposed model well merges the superpixels belonging to the same object into a region but the other method fails to do this and has severe segmentation leakage. Therefore, such examples prove that our models are robust not only in images taken under great exposures, but in those low-light and blurry scenarios. Additionally, we present simulations of an aircraft engine in Fig. 8c, which is difficult for another algorithm to perform accurate segmentation on the boundary of engine itself since the background information is similar to the engine. Nevertheless, our proposed method can generate compact segmentation without losing edge information.

12   ACCV-20 submission ID 938

Table 5. Result of adopting flipping technique.

Encoder Architecture	Flip	VI	PRI	SC
InceptionResNet	No	1.446	0.863	0.634
InceptionResNet	YES	1.411	0.864	0.647

5 Conclusion

In this study, a novel image segmentation algorithm that integrates the merit of the Fully Convolutional Network and superpixel-based segmentation is proposed. We call it the *DMMSS-FCN*.

With the use of superpixel, which allows us to be able to solve the generic image segmentation by highly efficient networks. That is, we covert the merging decision into a boundary keeping problem across all superpixel pairs. Hence, with the use of Fully Convolutional Networks, we can solve all those boundary keeping problems with just one forward pass. As a result, the proposed *DMMSS-FCN* algorithm is not only vastly faster than the other state-of-the-art algorithms, but also has high score in accuracy close to human performance.

Additionally, one can obtain large quantity of training data by varying the parameters of existing superpixel algorithms to produce various superpixel boundary maps as input training data. That is, our method does not require a lot of human-annotated ground truth to be fully trained.

In this study, we proposed a simple but effective deep-learning-based generic image segmentation algorithm that leverages the FCN for learning generic image segmentation. The proposed algorithm is very effective and simulations show that it is capable of producing reliable segmentation results under many circumstances. Since our algorithm do not require any post-processing technique and further information from user, the simplicity makes it extremely efficient and reliable for any generic image segmentation task. Therefore, we hope this work could offer a great idea to be implemented in the downstream tasks, since segmentation is a quite useful technique in many computer vision applications.
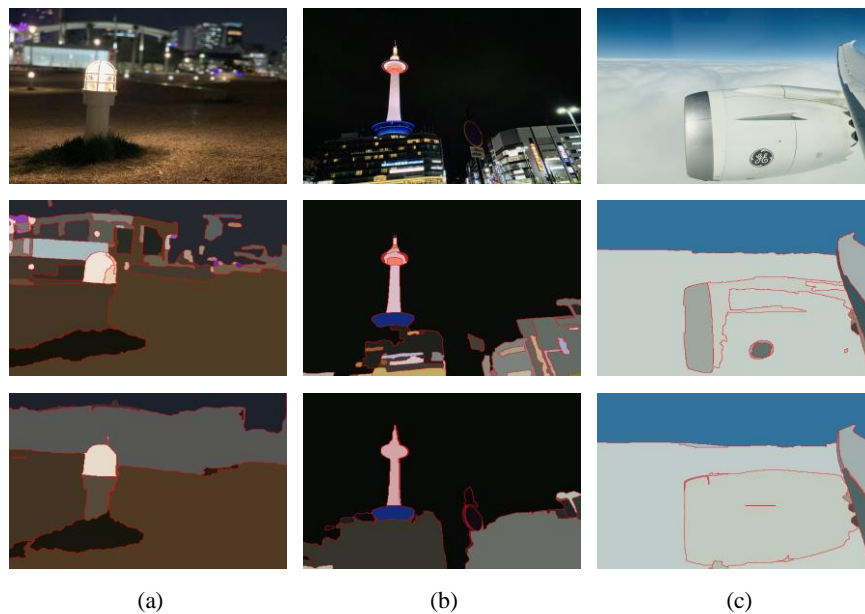
ACCV-20 submission ID 938   13

(a)   (b)   (c)

Fig. 8. Real-world image segmentation results. (1st row): original images. Results produced by (2nd row): gPb-OWT-UCM [4]; (3rd row): *DMMSS-FCN(ours)*.

ACCV2020
#938

ACCV2020
#938

CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.
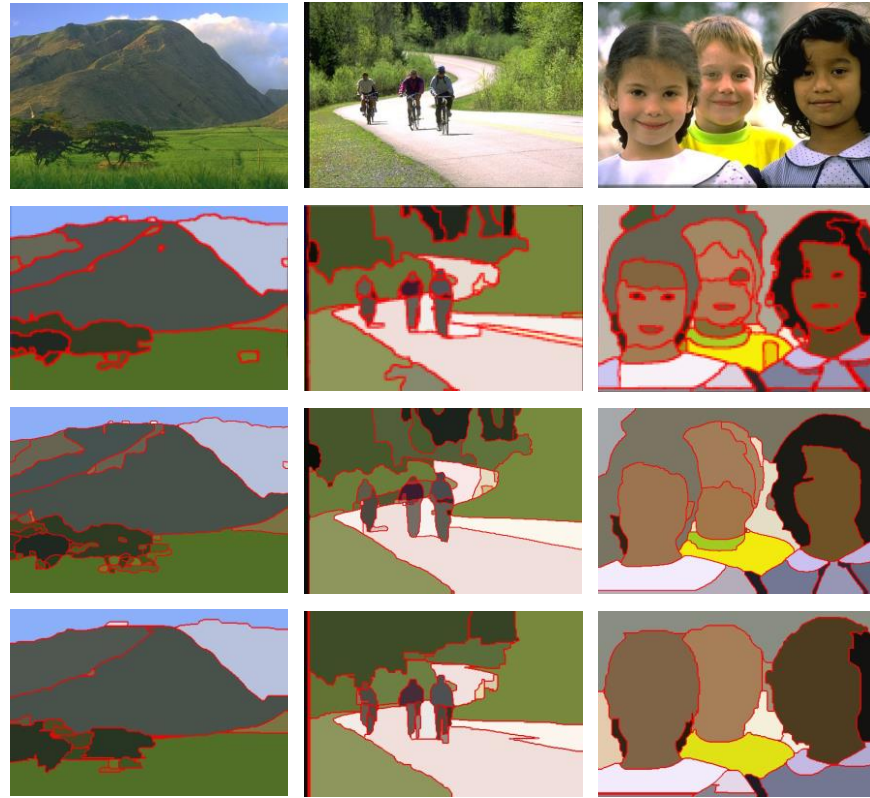
14   ACCV-20 submission ID 938

Fig. 9. Visual comparison between DC-Seg-full, gPb-OWT-UCM, and the proposed *DMMSS-FCN*. (1st row): original images. Results produced by (2nd row): DC-Seg-full [23]; (3rd row): gPb-OWT-UCM [4]; (4th row): *DMMSS-FCN(ours)*.

ACCV2020
#938

ACCV2020
#938

CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

ACCV-20 submission ID 938   15

References

1. Shih, F.Y., Cheng, S.: Automatic seeded region growing for color image segmentation. Image and Vision Computing **23** (2005) 877–886

2. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. **24** (2002) 603–619

3. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: From contours to regions: An empirical evaluation. In: IEEE Conf. Computer Vision and Pattern Recognition. (2009) 2294–2301

4. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **33** (2010) 898–916

5. Cour, T., Benezit, F., Shi, J.: Spectral segmentation with multiscale graph decomposition. In: IEEE Conf. Computer Vision and Pattern Recognition. **2** (2005) 1124–1131

6. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **22** (2000) 888–905

7. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. Int. J. Computer Vision **59** (2004) 167–181

8. Li, Z., Wu, X.M., Chang, S.F.: Segmentation using superpixels: A bipartite graph partitioning approach. In: IEEE Conf. Computer Vision and Pattern Recognition. (2012) 789–796

9. Kim, T.H., Lee, K.M., Lee, S.U.: Learning full pairwise affinities for spectral segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **35** (2012) 1690–1703

10. Yang, Y., Wang, Y., Xue, X.: A novel spectral clustering method with superpixels for image segmentation. Optik **127** (2016) 161–167

11. Xia, X., Kulis, B.: W-net: A deep model for fully unsupervised image segmentation. arXiv preprint arXiv:1711.08506 (2017)

12. Noh, H., Hong, S., Han, B.: Learning deconvolution network for semantic segmentation. In: IEEE Int. Conf. Computer Vision. (2015) 1520–1528

13. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: IEEE Conf. Computer Vision and Pattern Recognition. (2015) 3431–3440

14. 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. IEEE Trans. Pattern Anal. Mach. Intell. **40** (2017) 834–848

15. Liu, M.Y., Tuzel, O., Ramalingam, S., Chellappa, R.: Entropy rate superpixel segmentation. In: IEEE Conf. Computer Vision and Pattern Recognition. (2011) 2097–2104

ACCV2020
#938

ACCV2020
#938

CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

16   ACCV-20 submission ID 938

16. 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. **34** (2012) 2274–2282

17. Jampani, V., Sun, D., Liu, M.Y., Yang, M.H., Kautz, J.: Superpixel sampling networks. In: European Conf. Computer Vision. (2018) 352–368

18. Tu, W.C., Liu, M.Y., Jampani, V., Sun, D., Chien, S.Y., Yang, M.H., Kautz, J.: Learning superpixels with segmentation-aware affinity loss. In: IEEE Conf. Computer Vision and Pattern Recognition. (2018) 568–576

19. Haehn, D., Kaynig, V., Tompkin, J., Lichtman, J.W., Pfister, H.: Guided proofreading of automatic segmentations for connectomics. In: IEEE Conf. Computer Vision and Pattern Recognition (2018) 9319-9328

20. Kelm, A.P., Rao, V.S., Zolzer, U.: Object contour and edge detection with refinecontournet. In: Int. Conf. Computer Analysis of Images and Patterns. (2019) 246–258

21. Chen, L.C., Papandreou, G., Schroff, F., Adam, H.: Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587 (2017)

22. Szegedy, C., Loffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-ResNet and the impact of residual connections on learning. In: AAAI Conf. Artificial Intelligence. (2017) 4278-4284

23. Donoser, M., Schmalstieg, D.: Discrete-continuous gradient orientation estimation for faster image segmentation. In: IEEE Conf. Computer Vision and Pattern Recognition. (2014) 3158–3165

24. Taylor, C.J.: Towards fast and accurate segmentation. In: IEEE Conf. Computer Vision and Pattern Recognition. (2013) 1916–1922

25. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conf. Computer Vision and Pattern Recognition. (2016) 770–778

26. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: IEEE Conf. Computer Vision and Pattern Recognition. (2017) 1251–1258