


Article

Contour Detection for Fibre of Preserved Szechuan Pickle Based on Dilated Convolution

Hongyang Li ^{1,2} , Lizhuang Liu ^{1,*}, Zhenqi Han ¹ and Dan Zhao ¹¹ Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China² University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: liulz@sari.ac.cn

Received: 26 April 2019; Accepted: 27 June 2019; Published: 1 July 2019



Featured Application: Our method is applied to the automatic detection of the fibre of preserved Szechuan pickle. This method can be extended to the detection of non-salient contours and provide ideas for the detection of special contours.

Abstract: Peeling fibre is an indispensable process in the production of preserved Szechuan pickle, the accuracy of which can significantly influence the quality of the products, and thus the contour method of fibre detection, as a core algorithm of the automatic peeling device, is studied. The fibre contour is a kind of non-salient contour, characterized by big intra-class differences and small inter-class differences, meaning that the feature of the contour is not discriminative. The method called dilated-holistically-nested edge detection (Dilated-HED) is proposed to detect the fibre contour, which is built based on the HED network and dilated convolution. The experimental results for our dataset show that the Pixel Accuracy (PA) is 99.52% and the Mean Intersection over Union (MIoU) is 49.99%, achieving state-of-the-art performance.

Keywords: fibre of preserved Szechuan pickle; contour detection; dilated convolutions; HED

1. Introduction

The production of preserved Szechuan pickle includes peeling fibre, cleaning, cutting and so on. The cleaning and cutting processes can be easily and efficiently implemented with automatic devices, but peeling is always done manually, which limits the yield and quality of preserved Szechuan pickle. An automatic peeling device is needed to effectively solve this problem. Contour detection, as an important part of the device, is a necessary process to identify the location of fibre and guide the cutting tool.

In this paper, we address the problem of detecting contour in the fibre of preserved Szechuan pickle. The preserved Szechuan pickles consist of fibre, flesh and peel, as shown in Figure 1. The fibre of preserved Szechuan pickle is inedible, while the flesh can be eaten as a dish. In the productive process, the existence of fibre is not allowed and it is necessary to peel off the fibre.

By analyzing the section of the preserved Szechuan pickle stem, the contour of the preserved Szechuan pickle can be seen as one kind of non-salient contour. It is unsmoothed, irregular and the gray distinction between the fibre and flesh is small. The characteristics are not obvious enough and there is a large difference between individuals. The contour we studied in this paper is non-salient, which means that it is difficult to distinguish the contour from the background in texture, color and other aspects. This special contour also puts forward higher requirements for detection.

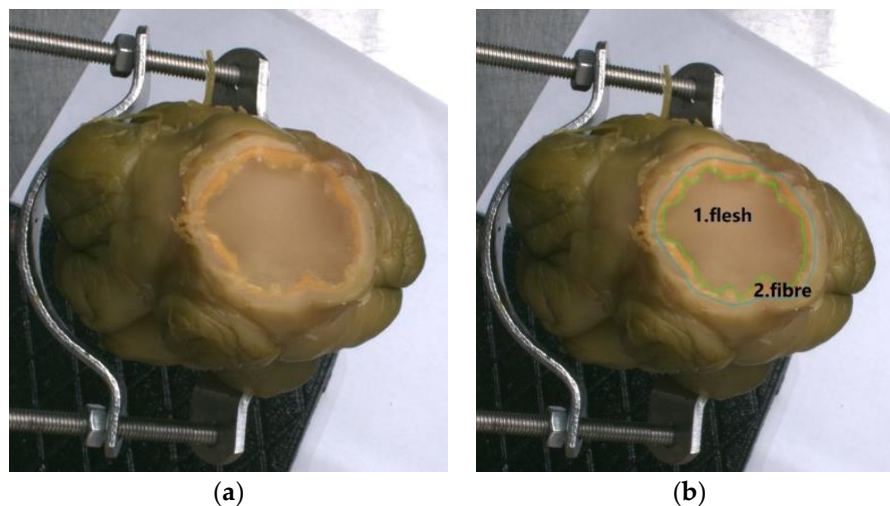


Figure 1. (a) Picture of a section of preserved Szechuan pickle; (b) the region within the green line is flesh, and the region surrounded by the blue and green line is fibre.

Contour detection has broad application prospects and is an important part of object segmentation, target detection, recognition and tracking [1]. There is a trend of using convolutional networks to detect contour since convolution neural networks show a strong capability to learn high-level representations of images. The effective integration of these methods and machines has promoted great developments in detection, such as the fully convolutional network (FCN) [2], holistically-nested edge detection (HED) [3], richer convolutional features for edge detection (RCF) [4] and so on. The features of the contour have the characteristics of large intra-class differences and small differences between classes. The general contour detection method has limited ability for these features. At present, there is little research into non-salient contour, but it is also one of the necessary technical means for production.

For this contour, we improve the structure of the HED network and use the output of each stage effectively. Our method automatically learns rich hierarchical representations and is able to make multi-scale predictions. We use dilated convolutions [5] to increase the receptive field. It is useful to reduce the loss of information and enhance the acquisition of spatial hierarchical information to improve the detection effect of the non-salient contours. In a complex background, our method can detect the region of contour. Finally, the Pixel Accuracy (PA) of our method [2] is 99.52% and the mean intersection over union (MIoU) [2] is 49.99%. Compared with the HED and RCF, the PA of our results increased by 3.80% and 9.61%, and the MIoU increased by 1.91% and 4.83% respectively.

2. Related Work

Contour detection has broad application prospects in computer vision, medical image and industrial production. At present, the main detection methods include the shallow feature-based method and deep feature-based method.

Shallow feature-based methods can be divided into edge-based, pixel-based and local region-based methods. Edge-based approaches are based on contour related edges or curves provided by edge detectors or human prior experience, aiming to determine whether they are contained in a certain contour [1]. Traditional operators in edge detection are widely used for high efficiency and strong applicability, such as Sobel, Laplace and Canny. Sobel [6] is a typical edge detection operator based on first derivative. Because it introduces a local average operation and has a smooth effect on noise and can eliminate the influence of noise very well. Laplace [7] is an isotropic operator and a second-order differential operator, which responds more strongly to isolated pixels than to edges or lines, so it is only suitable for noiseless images. Canny [8,9] is a multi-stage optimization operator with filtering, enhancement and detection. Its function is better than the previous examples. In pixel-based approaches, features are constructed and then employed to determine whether each pixel of the image

belongs to a contour [1]. The following three methods are pixel-based contour detection methods. Pb [10] is a probabilistic detector that combines discontinuity features with other gradient features, including color and texture gradients. Sparse code gradient (SCG) features can automatically learn from image data through sparse coding, thus minimizing human involvement [1,11]. In view of this point, an improved normalized cuts algorithm called multiscale combinatorial grouping (MCG) is proposed, providing a 20-fold increase in speed to the eigenvector computation [1,12]. Regarding contours as boundaries of interesting regions, region-based approaches take advantage of internal information of the regions to enhance their effectiveness and robustness [1]. The oriented watershed transform (OWT) was proposed by Jones et al. [13] and can form initial regions for the construction of an ultrametric contour map (UCM) [1,14,15], which also belongs to the region-based approaches.

Convolutional neural networks can extract deep features, which refer to the high-level representation of images. We will detail the method based on a convolutional neural network that is the most effective and suitable for a large number of samples. Deep convolutional neural networks (DCNNs) have recently shown impressive performance in various tasks such as classification, image and video detection, and segmentation [1]. AlexNet [16] is designed by Hinton and Alex Krizhevsky and it uses a GPU to speed up operations. ReLU, Dropout and LRN have been successfully applied in CNN for the first time. VGG net (Visual Geometry Group Network) [17] is a deep convolution neural network, which has 16 layers or 19 layers. In the structure, 3 × 3 filters which can effectively extract image features were used completely. The filters of GoogLeNet [18] have multiple scales, which solve the limitation of depth and width by stacking modules together. Jonathan Long et al. proposed FCN [2] for image segmentation. It converts the fully-connected layers of VGG into convolutional ones and attempts to harness information from multiple layers to better estimate the object boundaries [1]. The conditional random field (CRF) proposed by Lafferty et al. combines the characteristics of the maximum entropy model and hidden Markov model. It is also an undirected graph model [19]. In the same FCN framework, the dilated convolution is used to obtain more information about features, and the fully connected CRF is used to refine the label maps. The net is called DeepLab [20] and it can produce high-resolution segmentation. HED is an end-to-end approach based on FCN and VGG [4]. The edges of different scales are output through multiple side outputs, and the final edge output is obtained through a training weighted-fusion layer. HED improves the accuracy of edge detection through feature fusion. Inspired by the HED network, RCF was proposed in 2017 and achieved state-of-the-art performance on several available datasets [4]. It is very helpful to use this rich hierarchical information at each stage, so the model increases the number of output layers on the basis of HED.

With the development of contour detection technology, it is no longer difficult to recognize the general contour. Because of the non-saliency of the contour, many technical means cannot achieve the expected results. Under this condition, the detection of non-salient contour can be studied in depth on the basis of the existing contour detection technology and look for a technical means to distinguish this non-salient feature to form a complete and accurate contour.

3. Contour Detection for Fibre of Preserved Szechuan Pickle

3.1. HED Architecture

HED is improved and adjusted based on VGG16 net, and it shows a good performance in various tasks, containing 13 convolution layers and five side-outputs. The side-outputs contain multi-scale features extracted by the network. Supposing we have M side-output layers in the network, the classifier corresponding to each side-output can be defined as

$$L_{side}(W, w) = \sum_{m=1}^M \alpha_m l_{side}^{(m)}(W, w^{(m)}), \quad (1)$$

where L_{side} denotes the image-level loss function of the side-output, W denotes the set of all standard network layer parameters, the corresponding weight can be denoted as w and M is the number of layers of the side-output [3].

According to the position of side-outputs, the structure of HED can be divided into five stages. Multiple stages with different strides can capture the inherent scales of contour maps [3]. After inputting pictures, we obtain the contour map predictions from both the side-output layers and the weighted-fusion layer:

$$(\hat{Y}_{HED}) = \text{Average}(\text{CNN}(X, (W, w, h)^*)), \tag{2}$$

where $\text{CNN}(\cdot)$ denotes the map produced by the network, \hat{Y}_i denotes the predictions through i stage, X denotes the raw input image and h is the fusion weight.

This multi-scale and multi-level feature information is conducive to the transmission of holistic information and helps the network to obtain better prediction results. This structure is beneficial to our method. In the process of testing, the output results also prove that our choice is effective.

3.2. Dilated Convolutions

Dilated convolution increases the reception field by injecting holes into standard convolution maps. Compared with the original normal convolution, dilated convolution has a hyper-parameter called the dilation rate, which refers to the number of kernels intervals (e.g., the normal convolution is dilatation rate 1). The discrete convolution operator $*_l$ can be defined as

$$(F *_l k)(p) = \sum_{s+lt=p} F(s)k(t), \tag{3}$$

where F denotes a discrete function, and k is a discrete filter of size $(2r + 1)^2$. We will refer to $*_l$ as a dilated convolution [5].

Dilated convolution can solve problems such as loss of data structure, loss of spatial hierarchical information, and an inability to reconstruct small object information. For the contour of preserved Szechuan pickle, dilated convolution has a better effect on information extraction when the difference of feature information is small. We hope to give full rein to its advantages and it will play an active role in our research.

3.3. Dilated-HED

Inspired by many proposed contour detection models, we designed our own model for the contour detection of preserved Szechuan pickle on the basis of HED. In our repeated tests, this structure has the best result, as can be seen in Figure 2.

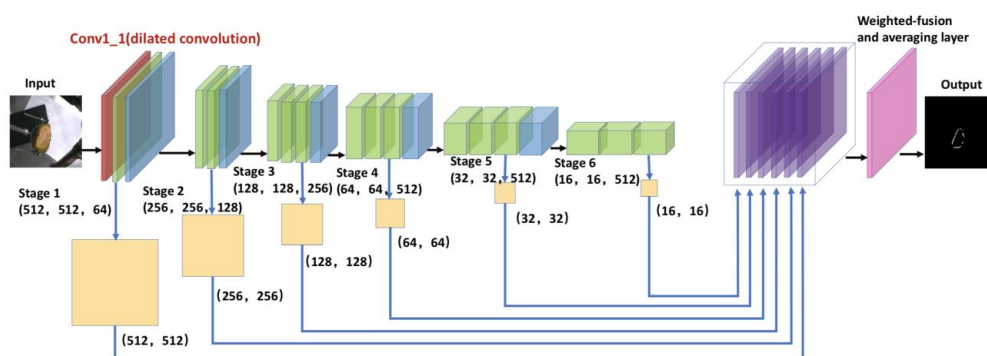


Figure 2. The structure of our model. The dark red cube represents the dilated convolution layer, the green cubes represent the convolution layer, the blue cubes represent the pooling layer, and the side-output layers are yellow. The purple cubes in the transparent cube represent the fusion of the output of each layer, and finally, a weighted-fusion layer is formed, represented by pink.

The input is an image and our network outputs a contour possibility map of the same size. Our model is improved on the basis of HED, adding a stage and dilated convolution, which is more sensitive to features. There are 16 convolution layers and five max-pooling layers. Its convolution layers are divided into six stages, in which a side-output layer is inserted after the last convolution of each stage. There are two convolution layers and one max-pooling layer in the first and the second stage. From the third to the fifth stage, each stage contains three convolution layers and one max-pooling layer. The last stage contains only three convolution layers. The outputs of dilated-HED are multi-scale and are finally integrated into a weighted-fusion layer by deconvolution. The weighted-fusion layer can automatically learn how to combine and average outputs from multiple scales, which are scaled according to 1, 2, 4, 8, 16 and 32. Our network architecture has different receptive field sizes and it will become larger with the deepening of the network. The convolution kernel of 3×3 is used in the HED network, and its receptive field is small. Increasing receptive field can capture more regional features. We use dilated convolution, which is located on the first layer of the whole structure. Our experimental results for the test data will be presented in the next section.

4. Experiments

4.1. Dataset of the Contour of Fibre

Due to the lack of data, we built a fibre contour dataset of preserved Szechuan pickle which was collected from production workshop. The object is the section of the fibre of preserved Szechuan pickle. We used a 5-million-pixel industrial camera with a 16 mm fixed-focus lens. The distance between the section of preserved Szechuan pickle and the camera is more than 30 cm to simulate the actual production process.

The data collected are in two forms: pictures and videos. Pictures serve as the main body of the dataset and videos serve as data supplements. Labels of each image are quality-controlled and human-annotated. The collected data are labeled manually by LabelMe [21], which is an open source annotation tool. We labeled images with two colors: black and white. The region of contour is white and the background is black. The dataset contains 2120 pairs of pictures. Each pair of pictures has an original picture and a label. It contains about 600 different individuals, with an average of approximately 3 images per stem. Some examples of the dataset are shown in Figure 3.

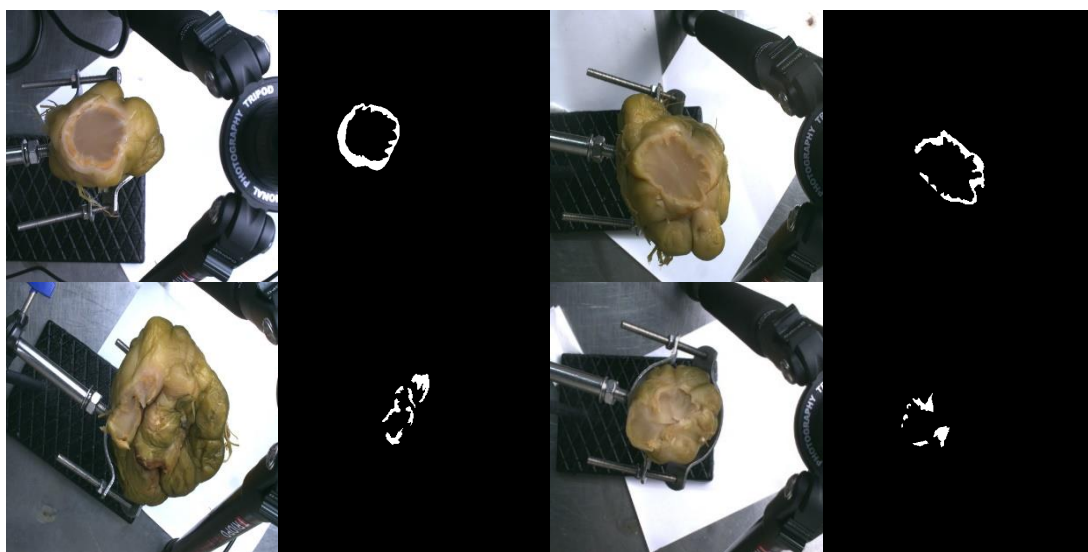


Figure 3. Some examples of the dataset. The two figures are a pair of data, with the image on the left and label on the right.

The method of marking a region can reduce the inaccuracy of contour lines and creates more possibilities to extract information from the original picture. We divide the dataset into the train set and test set in the ratio of 8:2, which comprise 1696 pairs and 424 pairs, respectively. The fibres of preserved Szechuan pickle are irregular. Some of them have contours that are continuous, while others are discrete regions. A continuous contour means that the contour consists of only one connected domain, while a discrete contour consists of two or more connected domains. The shape of the region of the continuous contour is similar to a ring, except that the inner and outer lines of the ring are unsmooth and irregular. The shape of the region of a discrete contour is usually a part of the ring and is mixed with many irregular shapes. The number of cases is shown in Table 1.

Table 1. Introduction to our dataset.

Categories	Continuous contour	Discrete contour	Total
Train set	910	786	1696
Test set	256	168	424

The fibre contour dataset of preserved Szechuan pickle is quite different from the public datasets, such as BSD500, PASCAL VOC2012 which are mainly for people, roads, vehicles and so on. The dataset of preserved Szechuan pickle is an object with small inter-class differences and large intra-class differences, and public datasets are the opposite. There are fewer common characteristics of objects between the public datasets and ours. It is necessary to set up a non-salient contour dataset and our method is also designed for this kind of dataset.

4.2. Comparison of RCF, HED and Dilated-HED

We evaluate dilated-HED, HED and RCF on the test set, which is composed of 424 pairs of pictures. The detection accuracy is evaluated using two measures: Pixel Accuracy (PA) and Mean Intersection over Union (MIoU). Both are proposed in [2] and are standard measures for semantic segmentation. Pixel accuracy (PA) is the simplest indicator used to calculate the ratio between the number of pixels correctly classified and the total number of pixels. It can be defined as

$$PA = \frac{\sum_{i=0}^k p_{ii}}{\sum_{i=0}^k \sum_{j=0}^k p_{ij}}, \quad (4)$$

$k+1$ represents the number of categories of contours, including background. p_{ij} represents the number of pixels that belong to class i but are predicted to be class j . MIoU is a standard metric to calculate the ratio of intersection and union between sets. Its definition is as follows:

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}}, \quad (5)$$

Before we loaded the data into the model for training, we first adjusted the size of the picture to 512×512 . The main reason for this is memory constraints and control parameters. The output images on the dataset are grayscale, with gray values ranging from 0 to 255. The value of 0 represents the background, and the value of pixel is closer to 255, the more likely it is to be a contour. By setting the threshold, we can filter out the possible points to get the predicted contours.

Compared with HED and RCF networks, our method achieves better processing results. For a specific comparison, see Figure 4.

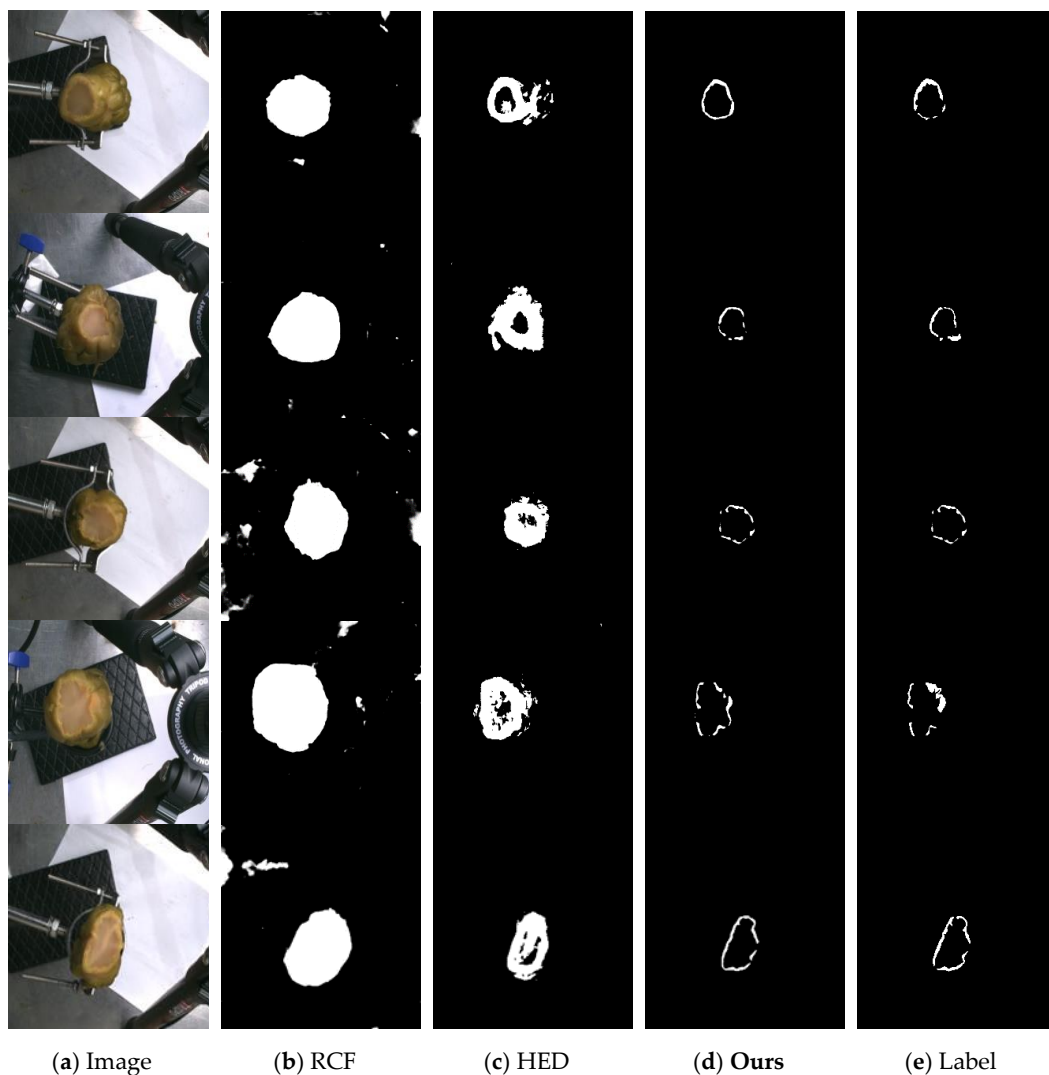


Figure 4. Comparison of richer convolutional features for edge detection (RCF), holistically-nested edge detection (HED) and dilated-HED. From left to right: (a) original images; (b) the results of RCF; (c) the results of HED; (d) the results of our method; (e) labels, also known as ground-truth.

From the comparison of the above results, we can see that RCF outputs more false regions, which are distributed on both sides of the right contour, and the background area of non-preserved Szechuan pickle is also recognized as a contour. HED can detect the area where the section of preserved Szechuan pickle is located, but there are also many false regions. RCF and HED have low accuracy and high misjudgment rates, which means they are not suitable for contour detection in our project. Our model can detect the contour of fibre accurately, which is distributed irregularly. In addition, the PA and MIoU of the three methods are measured as shown in Table 2.

Table 2. Pixel Accuracy (PA) and Mean Intersection over Union (MIoU) by three methods (Unit: %).

Categories	RCF	HED	Ours	Continuous Contour (Ours)	Discrete Contour (Ours)
PA	89.91	95.72	99.52	99.52	99.53
MIoU	45.16	48.08	49.99	49.99	49.98

In the evaluation index, the background is also a category, and the background is much larger than the target in the image, so the value of PA is large no matter which method is used. The result shows that PA of our method is 99.52% and MIoU is 49.99%. We use dilated convolution at conv1-1 and the dilated rate is [(2,2), (1,1)]. From the data, we can see that the values of PA and MIoU obtained by

RCF are the lowest, followed by HED, and our model gets the best results. This is consistent with the conclusion obtained from Figure 3. Compared with the HED and RCF, the PA of our results increased by 3.80% and 9.61%, and the MIoU increased by 1.91% and 4.83% respectively. Dilated-HED is more effective in the detection of fibre contours as shown in the experiment.

We divided the test set into two sets according to the contour types, and then use our proposed model to test on two sets of data. The results in Table 2 show that the PA of the discrete contour is higher than the continuous contour by about 0.01%, but the MIoU of the discrete contour is lower than the continuous contour by about 0.01%. By analyzing the data and predicted images, we believe that our model will smoothen regions with large variation, which has a greater impact on the continuous contour. So, the PA of the two sets will be different. The model is more effective for large contours and the smaller connected areas in discrete contours are more easily omitted, so the MIoU of continuous contours is higher.

The network structure of HED improves the performance of the model by incorporating multi-scale information. We retain this structure and add the number of layers of the model, so that the model can get more useful information. When we obtain the feature map, it is beneficial to use dilated convolution first in our model. Dilated convolution enlarges the receptive field and is more sensitive to non-salient features. This extraction method can capture the key features of contour, so it can make effective use of non-salient features. From the results, we can see that these changes have a promotion effect.

4.3. Training Details

In this part we discuss our detailed implementation. We experimented with our method and HED in the Tensorflow [22] deep learning framework and trained it on a Nvidia GTX 1080-TI GPU. RCF was implemented in the PyTorch [23] deep learning framework and trained on the same GPU. Like HED, our model has filters with size 3×3 . The pool size of the pooling layer is 2×2 and the strides is 2×2 . We initialized the network with the weight of VGG16 training on ImageNet ILSVRC—2014 submission [17]. To get a better optimization, we changed the initial values of the parameters in the convolution layer. The initial values of the parameters are different at different layers. An Adam optimizer with a learning rate of 0.0001 is used in our method. The batch size is 2, the number of decay_steps is 10000 and decay_rate is 0.1. We set pos_weights at 0.7, weight_decay_ratio at 0.0002 and sides_weights at 1.0 in every output-layer. It takes about 5 h to train our model.

5. Discussion and Future Work

In this paper, we proposed a dilated convolutional network for non-salient contour detection. Our method shows promising results in performing image-to-image learning by combining dilated convolution and multi-scale visual responses. Dilated-HED takes advantage of HED in contour recognition, combines dilated convolution and makes the network more sensitive to the fibre of preserved Szechuan pickle by enhancing the ability to acquire non-salient features. In order to achieve better results, we will try to extract the section of preserved Szechuan pickle and then detect the contour. In the case of large background and small target, this method can reduce the interference of background to accuracy. We believe that the work is a step towards non-salient contour detection which is a crucial step in industrial production. Non-salient contour detection will greatly improve the accuracy of production and make products more sophisticated. We will also add the categories of contours and make our method applicable to more fields in the future.

Author Contributions: All the authors contributed to this work. Conceptualization, H.L. and Z.H.; Data curation, H.L.; Formal analysis, H.L. and Z.H.; Investigation, H.L. and D.Z.; Methodology, H.L. and L.L.; Project administration, L.L.; Validation and visualization, H.L.; Writing—original draft, H.L.; Writing—review & editing, L.L. and Z.H.

Funding: This research was funded in part by grants from the Science and Technology Commission of Shanghai Municipality (17511106400, 17511106402).

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

References

1. Gong, X.Y.; Su, H.; Xu, D.; Zhang, Z.T.; Shen, F.; Yang, H.B. An Overview of Contour Detection Approaches. *Int. J. Autom. Comput.* **2018**, *15*, 656–672. [CrossRef]
2. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [CrossRef]
3. Xie, S.; Tu, Z. Holistically-Nested Edge Detection. *Int. J. Comput. Vis.* **2017**, *125*, 3–18. [CrossRef]
4. Liu, Y.; Cheng, M.M.; Hu, X.; Wang, K.; Bai, X. Richer Convolutional Features for Edge Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**. Available online: <http://mmcheng.net/rcfEdge/> (accessed on 29 January 2019).
5. Yu, F.; Koltun, V. Multi-scale context aggregation by dilated convolutions. *arXiv* **2016**, arXiv:1511.07122.
6. Kanopoulos, N.; Vasanthavada, N.; Baker, R.L. Design of an image edge detection filter using the Sobel operator. *IEEE J. Solid-State Circuits* **2002**, *23*, 358–367. [CrossRef]
7. Van Vliet, L.J.; Young, I.T.; Beckers, G.L. A nonlinear laplace operator as edge detector in noisy images. *Comput. Vis. Gr. Image Process.* **1989**, *45*, 167–195. [CrossRef]
8. Ding, L.; Goshtasby, A. On the Canny edge detector. *Pattern Recognit.* **2001**, *34*, 721–725. [CrossRef]
9. Bao, P.; Zhang, L.; Wu, X. Canny Edge Detection Enhancement by Scale Multiplication. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1485–1490. [CrossRef] [PubMed]
10. Martin, D.R.; Fowlkes, C.C.; Malik, J. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 530–549. [CrossRef] [PubMed]
11. Ren, X.F.; Bo, L. Discriminatively trained sparse code gradients for contour detection. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 584–592.
12. PArbeláez, P.; Pont-Tuset, J.; Barron, J.T.; Marques, F.; Malik, J. Multiscale combinatorial grouping. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 328–335. [CrossRef]
13. Jones, H.E.; Grieve, K.L.; Wang, W.; Sillito, A.M. Sillito. Surround suppression in primate V1. *J. Neurophysiol.* **2001**, *86*, 2011–2028. [CrossRef] [PubMed]
14. Arbelaez, P. Boundary extraction in natural images using ultrametric contour maps. In Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06), New York, NY, USA, 17–22 June 2006; pp. 182–182. [CrossRef]
15. Arbelaez, P.; Maire, M.; Fowlkes, C.; Malik, J. From contours to regions: An empirical evaluation. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 2294–2301. [CrossRef]
16. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems—Volume 1, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
17. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
18. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2014; Available online: <https://arxiv.org/abs/1409.4842> (accessed on 15 March 2018).
19. Lafferty, J.; McCallum, A.; Pereira, F.C.N. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In Proceedings of the 18th International Conference on Machine Learning 2001 (ICML 2001), San Francisco, CA, USA, 28 June–1 July 2001; pp. 282–289.
20. 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.* **2018**, *40*, 834–848. [CrossRef] [PubMed]
21. Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. LabelMe: A Database and Web-Based Tool for Image Annotation. *Int. J. Comput. Vis.* **2008**, *77*, 157–173. [CrossRef]

22. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A system for large-scale machine learning. In Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation, Savannah, GA, USA, 2–4 November 2016; Volume 16, pp. 265–283. Available online: <https://arxiv.org/abs/1605.08695v2> (accessed on 10 August 2018).
23. Ketkar, N. Introduction to PyTorch. In *Deep Learning with Python: A Hands-on Introduction*; Apress: Berkeley, CA, US, 2017; pp. 195–208.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).