Weakly Supervised Retinal Detachment Segmentation Using Deep Feature Propagation Learning in SD-OCT Images

Tieqiao Wang, Sijie Niu(✉), Jiwen Dong, and Yuehui Chen

Shandong Provincial Key Laboratory of Network Based Intelligent Computing,
School of Information Science and Engineering,
University of Jinan, Jinan 250022, China
sjniu@hotmail.com

Abstract. Most automated segmentation approaches for quantitative assessment of sub-retinal fluid regions rely heavily on retinal anatomy knowledge (e.g. layer segmentation) and pixel-level annotation, which requires excessive manual intervention and huge learning costs. In this paper, we propose a weakly supervised learning method for the quantitative analysis of lesion regions in spectral domain optical coherence tomography (SD-OCT) images. Specifically, we first obtain more accurate positioning through improved class activation mapping; second, in the feature propagation learning network, the multi-scale features learned by the slice-level classification are employed to expand its activation area and generate soft labels; finally, we use generated soft labels to train a fully supervised network for more robust results. The proposed method is evaluated on subjects from a dataset with 23 volumes for cross-validation experiments. The experimental results demonstrate that the proposed method can achieve encouraging segmentation accuracy comparable to strong supervision methods only utilizing image-level labels.

Keywords: SD-OCT images · Medical image segmentation · Weakly supervision learning · Convolutional neural network

1 Introduction

Central serous chorioretinopathy (CSC) is one of the most common retinopathies, which is common among middle-aged men [2]. Neurosensory retinal detachment (NRD) and pigment epithelial detachment (PED) are the main manifestations of CSC. With the development of retinal imaging techniques, SD-OCT becomes an increasingly popular method for diagnosing ophthalmic diseases with the ability to provide high-resolution, cross-sectional and three-dimensional representations. Therefore, measuring the quantity of CSC and monitoring its change over time is

This work was supported by the National Natural Science Foundation of China under Grant No. 61701192, 61671242, 61872419, 61873324.

H. Fu et al. (Eds.): OMIA 2020, LNCS 12069, pp. 146–154, 2020.
https://doi.org/10.1007/978-3-030-63419-3_15

of significant importance in clinical assessment. However, manual delineation is laborious even for experienced experts and often suffers from inter-variability.

Currently, to address the above problem, many methods have been proposed to quantify and analyze the CSC in SD-OCT images. Mathematical based methods (e.g. level sets) [7,13], graph search model [14], enface fundus driven method [15] have been proposed to segment the subretinal fluid. Later, semi-supervised approaches [3,16,17] have been presented to address the problem of low contrast and speckle noise in SD-OCT image. The supervised learning methods, including random forest [5], K nearest neighbor [8], kernel regression [11], and deep learning [12], have been introduced to extracting the fluid regions in SD-OCT image. With the drastic advance of deep learning, recent deep learning networks have demonstrated successful performance of image segmentation tasks. A RelayNet proposed by Roy et al. [10] is employed to obtain the retinal layers and fluid regions. Gao et al. [4] proposed a novel image-to-image double-branched and area-constraint fully convolutional networks (DA-FCN). Hrvoje Bogunovic et al. [1] introduced a benchmark and reported the analysis of the challenge RETOUCH. Although deep-learning methods have achieved existing results for segmenting CSC in SD-OCT images, they all rely heavily on manual pixel-level annotations [4] and some forms of complicated data preparations [16] like layer segmentation. These problems greatly increase labor costs and limit related applications.

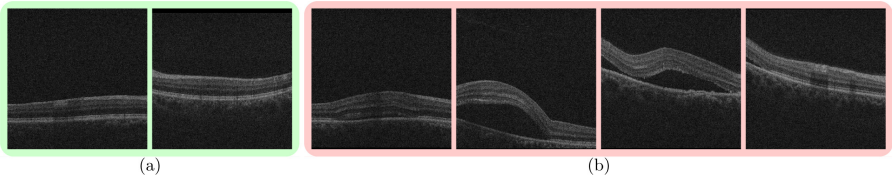

Fig. 1. The picture-level annotations used in our proposed method, normal (Fig. 1(a)) and abnormal slices (Fig. 1(b)) are labeled as 0 and 1, respectively.

To tackle the above challenging problems, weakly supervised learning methods are widely studied recently. Among various forms of weak annotations, image-level labels are the most user-friendly, which is convenient and cheap to collect. Unlike pixel-level labels and scribble annotations, which require more or less accurate outlines and locations of target areas, image-level labels only need to give a general category judgment for a specific slice. Thus, the image-level annotation method not only greatly improves the labeling efficiency of professional clinicians, but also makes labeling by the enthusiastic citizens possible. In our task, as the slices are shown in Fig. 1, the images with subretinal fluid regions are considered as abnormal and annotated 1, while the images without such lesion regions are taken as normal and labeled 0. To our best knowledge, there are no weakly supervised methods utilizing image-level labels to achieve CSC segmentation in OCT images. Therefore, to develop a weakly supervised

learning method using these available weakly labeled data is becoming significantly important.

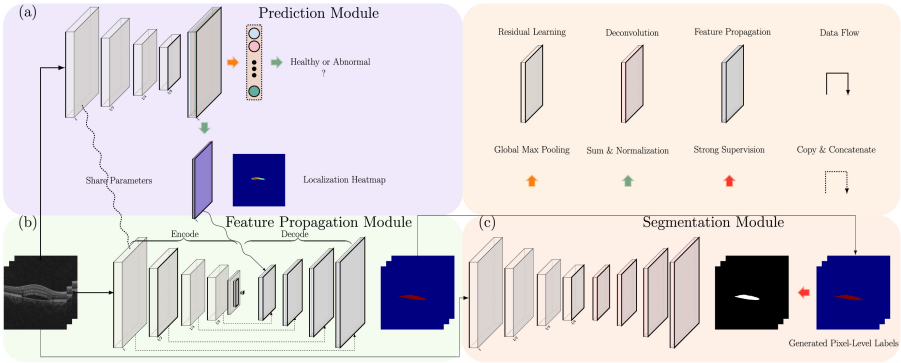

Fig. 2. The proposed weakly supervised network has three parts: (a) Prediction Module obtains a more pure activation region; (b) Feature Propagation Module benefits from features of different sizes of the network to expand the influence of the activation region; and (c) Segmentation Module uses feature propagation result to train and obtain more robust segmentation.

In this paper, we propose a weakly supervised deep feature propagation learning for retinal detachment segmentation in SD-OCT images using subjects with image-level annotations. The schematic diagram of our method is shown in Fig. 2. The proposed method has the following three contributions: (1) The class activation mapping was improved to achieve accurate localization inside the lesion regions using only the slice category labels. (2) An efficient feature propagation model was proposed, which utilizes deep to shallow features learned by the classification network to expand the impact of the active area. (3) The most efficient labeling in the medical field is the 0–1 binary judgment of health or abnormality. The propagation learning method employs only slice categories and achieves a segmentation effect comparable to strong pixel-level supervision, which is of far-reaching significance in the field of computer-assisted medical image analysis by reducing annotation costs and improving learning efficiency.

2 Methodology

The proposed feature propagation learning method consists of three modules, as the architecture overview shown in Fig. 2. The prediction module uses a fully convolutional network similar to the class activation mapping (CAM) [18], which learns to predict the saliency map of the lesion region from the input image. The feature propagation module propagates, constrains, and refines the salient features of the prediction module by propagating this feature among

multi-scale feature maps to generate pixel-level soft labels. The fully supervised module trains end-to-end strong supervised segmentation networks using soft labels generated by feature propagation learning.

In the following sections, we first describe the prediction module in Sect. 2.1, then the operating mechanism of our newly designed feature propagation learning module in Sect. 2.2. The fully supervised module is described in Sect. 2.3.

2.1 Saliency Map Generation Based on Improved CAM

Inspired by the CAM method [18], we utilize global pooling to activate salient lesion regions, since this strategy is able to capture general location information of lesion regions only using image-level labels. To make the positioning of the prediction module more accurate and located inside the target area, we use global maximum pooling (GMP) instead of global average pooling (GAP) to process the features generated by the last layer of convolution.

In the proposed class activation mapping module, we keep the input convolution layer and the first three stages of ResNet-50 except for the pooling operation after the input convolution. Then, we use global maximum pooling to output the spatial maximum of the feature map of each unit in the last convolution layer to train a binary classifier.

Specifically, let donate $\mathcal{F} \in \mathbb{R}^{C \times \frac{h}{8} \times \frac{w}{8}}$ as the feature maps of in the last convolution, where C, h, w represent channel, height, width. The maximum response obtained by using global maximum pooling is defined as:

$$f_c = \max \mathcal{F}_c, c \in \{1, 2, ..., C\}, \mathcal{F}_c \in \mathbb{R}^{\frac{h}{8} \times \frac{w}{8}} \tag{1}$$

In this task, slice identification (i.e., normal or abnormal) is taken as a binary classification and BCE loss is the most widely used loss in such tasks. So we use it as the loss function, which is defined as:

$$\mathcal{L}_{bce} = -\sum_{1}^{N} [g_n \log r_n + (1 - g_n) \log(1 - r_n)] \tag{2}$$

where N is the total number of training slices; for a specific slice, $g_n \in \{0, 1\}$ donates the ground truth label and $r_n \in (0, 1)$ represents the binary classification result of the prediction module, which is defined as:

$$r_n = \frac{1}{1 + \exp(-\sum_1^C f_c)} \tag{3}$$

As previously explained, f_c represents the global maximum of the feature map, reflecting the most significant part of the feature map in each of the C channels. We use the sum of these maximum values to generate 0 (normal) or 1 (abnormal) category results, which significantly enhances the network's ability to accurately obtain the location of the lesion. Because generating the network output in this way is helpful for activating (learning to form larger feature values) feature

maps containing significant abnormal features, while suppressing (learning to form smaller feature values) containing normal tissues or background. Next, we also employ the same strategy to obtain the salient activation map $\mathcal{M} \in \mathbb{R}^{\frac{h}{8} \times \frac{w}{8}}$:

$$\mathcal{M}(x, y) = \sum_{i=1}^{c} \mathcal{F}_i \tag{4}$$

Generally, both global maximum pooling and global average pooling can highlight the distinguishable regions beneficial for classification. The latter is more widely used since it can provide larger distinguishable regions. However, in our task, the global average pooling (Fig. 3(a)) mainly focuses on the retinal layer structure information rather than the lesion regions. Fortunately, the global maximum pooling (Fig. 3(b)) identifies the most distinguishable regions, which are the part of regions we want to segment. Therefore we utilize the global maximum pooling to determine the positioning of the lesion. This is of great significance for segmenting lesion regions with evenly distributed textures. Thanks to this characteristic, the classification network can obtain the accurate location information and the feature distribution of lesion regions simultaneously.

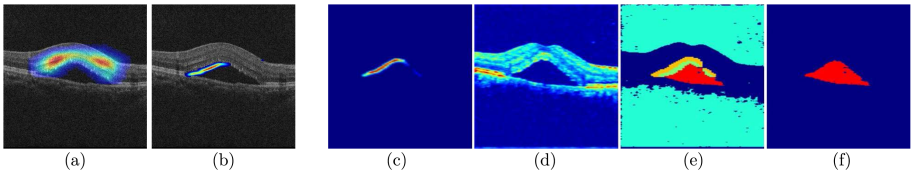

Fig. 3. The first two images show the comparison results using different pooling methods, where (a) and (b) respectively represent the global average pooling and the global maximum pooling; the last four images demonstrate the mechanism and results of feature propagation module, where figure (c), (d), (e), and (f) respectively represent the saliency map, reference map, intersection regions, and results obtained by feature propagation learning.

2.2 Soft Label Using Feature Propagation Learning

The feature propagation module (Fig. 2(b)) is a U-Net-like Encoder-Decoder network [9]. However, different from the decoder in strong supervised networks, we utilize multi-scale features generated by the classification network as guide information to propose a feature propagation learning module, thus this module involves no extra parameters.

In the encoder part, the architecture retains all the structures of ResNet-50 consisted of 5 residual block groups, where the input convolution layer shares weights with the prediction module. This part learns from the category information to generate multi-scale feature maps and then we propose a feature propagation module without learning. During the training process, the encoder part

will generate five saliency maps, which will be propagated in reverse order to the decoder as reference maps represented by $\mathcal{R}_i(i \in \{1, 2, 3, 4, 5\})$ (Fig. 2(b)).

The goal of the decoder part is to highlight the lesion regions using the proposed two feature propagation strategies, thereby generating optimized soft labels. Specifically, based on the characteristic that the prediction probability of the lesion regions is identical to that of the background, the mechanism of feature propagation in the decoder is defined as:

$$\mathcal{S}_{i-1} = \mathcal{R}_i \cap (\uplus \mathcal{S}_i), i \in \{4, 3, 2, 1\}, \mathcal{S}_4 = \mathcal{M} \tag{5}$$

where $\cap$ and $\uplus$ represent the saliency region intersection and the expansion operation. $\mathcal{S}_i$ represents the saliency map obtained at the i-th stage, where the initial localization heatmap is $\mathcal{M}$ generated by the prediction module, and $\mathcal{S}_0$ is the final segmentation result of feature propagation module.

During feature propagation learning (Fig. 3(c–f)), we first expand the saliency region obtained in the previous step, and then perform threshold segmentation on the saliency map and the reference map respectively. After the two images are superimposed, several different connectivity areas are generated, where the green part in Fig. 3(e) indicates the intersection of the two images. In the extended part, red regions in Fig. 3(e) indicates that the predicted probability value is similar to the intersection part, and yellow regions in Fig. 3(e) indicates the remaining part. In this way, the region similar to the intersection is preserved, and the segmentation results are iteratively updated and optimized after repeated feature propagations.

2.3 Lesion Segmentation with Strong Supervised Network

The soft labels generated by the feature propagation learning network train the fully supervised network (e.g., Deeplab V3, U-net) to obtain a more robust segmentation result. The problem solved here is that during the training process, a small sample of the localization network does not make correct classification, but it will have a more severe impact on the feature propagation learning. Therefore, we masked this problem with category information and trained a fully supervised segmentation network with correct generate labels.

3 Experiments

3.1 Datasets and Evaluation Metrics

Our analysis is based on one challenging dataset with NRD-fluid from [15,16], including 15 patients comprised of 23 vol (one patient may include more than one volume); each volume contains 128 images of 512 × 1024 pixels. This work was approved by the Institutional Review Board (IRB) of the First Affiliated Hospital of Nanjing Medical University with informed consent. Ground truth segmentation outlines were obtained by two experienced retinal experts.

The performance of segmentation accuracy is evaluated by the following metrics: 1) dice similarity coefficient (DSC), 2) true positive volume fraction (TPVF), and 3) positive predicative value (PPV). We denote R, TP, FP, and G as result of the method, true positive set, false positive set, and ground truth, respectively. These evaluation metrics are defined as: $DSC = \frac{2\times TP}{R+G}$, $TPVF = \frac{TP}{G}$, $PPV = \frac{TP}{TP+FP}$.

Table 1. The segmentation accuracy results obtained by different methods, two experts annotated the ground-truth labels (only for evaluation); where SoftGT (soft label ground truth) donates the segmentation results obtained from feature propagation module, Ours+U and Ours+D respectively represent the experimental results utilizing the fully supervised network U-net and Deeplab V3 with soft labels.

Method	Expert 1			Expert 2		
	DSC (%)	TPVF (%)	PPV (%)	DSC (%)	TPVF (%)	PPV (%)
LPHC [14]	65.7 ± 10.5	81.2 ± 9.3	55.8 ± 12.8	65.3 ± 10.4	81.3 ± 9.4	55.6 ± 13.3
FLSCV [13]	79.4 ± 20.2	84.4 ± 15.1	63.4 ± 7.3	78.9 ± 21.7	84.4 ± 16.0	86.2 ± 7.3
U-Net [9]	83.7 ± 6.2	90.3 ± 0.35	91.5 ± 6.8	84.3 ± 7.9	91.4 ± 5.2	92.8 ± 5.6
SS-KNN [8]	85.9 ± 4.1	80.3 ± 6.5	91.8 ± 3.8	86.1 ± 4.1	80.9 ± 6.6	92.3 ± 3.8
FCN [6]	87.0 ± 15.5	82.4 ± 19.9	95.9 ± 5.1	86.6 ± 15.6	82.6 ± 20.0	94.9 ± 5.1
RF [5]	88.9 ± 4.2	92.5 ± 4.3	91.9 ± 2.2	87.1 ± 4.3	92.6 ± 4.4	92.4 ± 2.0
CMF [16]	94.3 ± 2.6	92.0 ± 3.9	94.0 ± 3.5	93.9 ± 2.5	92.1 ± 4.1	93.0 ± 3.4
EFD [15]	94.6 ± 4.1	94.3 ± 5.1	94.1 ± 5.3	93.7 ± 4.0	94.2 ± 5.2	93.0 ± 4.8
DA-FCN [4]	**95.6 ± 1.6**	**94.4 ± 3.2**	**97.0 ± 1.1**	**95.0 ± 1.8**	**94.3 ± 1.8**	**95.8 ± 1.4**
SoftGT	86.9 ± 6.7	85.1 ± 11.5	90.0 ± 4.6	86.3 ± 6.7	84.9 ± 11.7	89.0 ± 4.7
Ours+U	91.2 ± 1.3	91.8 ± 2.1	90.7 ± 3.3	90.5 ± 1.2	91.5 ± 1.9	89.7 ± 3.0
Ours+D	92.4 ± 1.8	91.9 ± 2.3	93.0 ± 3.2	92.0 ± 2.0	91.8 ± 2.2	92.3 ± 3.4

3.2 Comparison Experiments

We compare our proposed model with nine state-of-the-art segmentation methods in SD-OCT images, including **(1) traditional segmentation methods:** label propagation and higher-order constraint (LPHC) [14], fuzzy level set with cross-sectional voting (FLSCV) [13], stratified sampling k-nearest neighbor classifier based algorithm (SS-KNN) [8], enface fundus-driven method (EFD) [15]; and **(2) strong supervised machine learning methods:** random forest classifier based method (RF) [5], continuous max-flow approach (CMF) [16], fully convolutional networks (FCN) [6], U-Net [9], and double-branched and area-constraint network (DA-FCN) [4].

Table 1 shows the mean DSC, TPVF, and PPV of all comparison methods for segmenting CSC. It can be seen that the performance of our proposed method is close to strong supervised methods. We illustrate the performance of different methods in several challenging and representative slices in Fig. 4. It can be seen that feature propagation learning has excellent segmentation potential to achieve excellent lesion boundary segmentation.

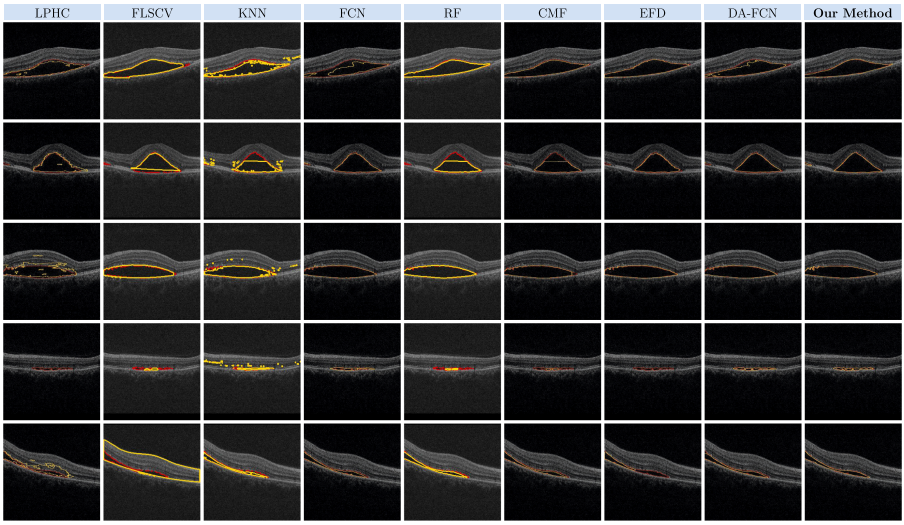

Fig. 4. Qualitative comparison of the proposed feature propagation learning method with other methods, where the yellow line represents the segmentation results of a certain method, and the red line represents the ground truth given by an ophthalmologist. (Color figure online)

4 Conclusion

In this paper, we present a weakly supervised method for segmenting CSC, in which only image-level annotation data collected is available for model training. To accurately locate the saliency region of the lesions, we improved the CAM to further focus on the targets. And also, we proposed an efficient feature propagation methods to generate the soft label for training segmentation model. Experiments demonstrate that our proposed method achieves a comparable effect to that of full-supervision and traditional techniques. Moreover, our proposed method is more straightforward and has cheaper implementation costs, which is a good inspiration for clinical research of retinal-related lesion segmentation in SD-OCT images.

References

1. Bogunović, H., et al.: RETOUCH: the retinal oct fluid detection and segmentation benchmark and challenge. IEEE Trans. Med. Imaging **38**(8), 1858–1874 (2019)
2. Dansingani, K.K., et al.: Annular lesions and catenary forms in chronic central serous chorioretinopathy. Am. J. Ophthalmol. **166**, 60–67 (2016)
3. Fernandez, D.C.: Delineating fluid-filled region boundaries in optical coherence tomography images of the retina. IEEE Trans. Med. Imaging **24**(8), 929–945 (2005)
4. Gao, K., et al.: Double-branched and area-constraint fully convolutional networks for automated serous retinal detachment segmentation in sd-oct images. Comput. Methods Programs Biomed. **176**, 69–80 (2019)

5. Lang, A., et al.: Automatic segmentation of microcystic macular edema in OCT. Biomed. Opt. Express **6**(1), 155–169 (2015)
6. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
7. Novosel, J., Wang, Z., de Jong, H., Van Velthoven, M., Vermeer, K.A., van Vliet, L.J.: Locally-adaptive loosely-coupled level sets for retinal layer and fluid segmentation in subjects with central serous retinopathy. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 702–705. IEEE (2016)
8. Quellec, G., Lee, K., Dolejsi, M., Garvin, M.K., Abramoff, M.D., Sonka, M.: Three-dimensional analysis of retinal layer texture: identification of fluid-filled regions in SD-OCT of the macula. IEEE Trans. Med. Imaging **29**(6), 1321–1330 (2010)
9. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
10. Roy, A.G., et al.: ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. Biomed. Opt. Express **8**(8), 3627–3642 (2017)
11. Schaap, M., et al.: Coronary lumen segmentation using graph cuts and robust kernel regression. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 528–539. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02498-6_44
12. Venhuizen, F.G., et al.: Deep learning approach for the detection and quantification of intraretinal cystoid fluid in multivendor optical coherence tomography. Biomed. Opt. Express **9**(4), 1545–1569 (2018)
13. Wang, J., et al.: Automated volumetric segmentation of retinal fluid on optical coherence tomography. Biomed. Opt. Express **7**(4), 1577–1589 (2016)
14. Wang, T., et al.: Label propagation and higher-order constraint-based segmentation of fluid-associated regions in retinal SD-OCT images. Inf. Sci. **358**, 92–111 (2016)
15. Wu, M., et al.: Automatic subretinal fluid segmentation of retinal SD-OCT images with neurosensory retinal detachment guided by enface fundus imaging. IEEE Trans. Biomed. Eng. **65**(1), 87–95 (2017)
16. Wu, M., et al.: Three-dimensional continuous max flow optimization-based serous retinal detachment segmentation in SD-OCT for central serous chorioretinopathy. Biomed. Opt. Express **8**(9), 4257–4274 (2017)
17. Zheng, Y., Sahni, J., Campa, C., Stangos, A.N., Raj, A., Harding, S.P.: Computerized assessment of intraretinal and subretinal fluid regions in spectral-domain optical coherence tomography images of the retina. Am. J. Ophthalmol. **155**(2), 277–286 (2013)
18. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2921–2929 (2016)