Partially Annotated Gastric Pathological Image Classification

Yanping Cui[1], Zhangcheng Wang[1], Guanzhen Yu[2(✉)], and Xinmei Tian[1(✉)]

[1] CAS Key Laboratory of Technology in Geo-Spatial Information Processing and Application Systems, University of Science and Technology of China, Hefei, China
{cuiyp,wzc1}@mail.ustc.edu.cn, xinmei@ustc.edu.cn

[2] Department of oncology, Longhua Hospital affiliated to Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China
qiaoshanqian@aliyun.com

Abstract. Previous works mainly address the medical datasets with image-wise labels or pixel-wise labels. However, it is difficult to train a model with only image-wise labels, and pixel-wise labels commonly refer to the high expense of annotations. A feasible solution is to make a compromise between data annotation and the performance. In this paper, we propose a cascaded convolutional neural network framework to classify partially annotated pathological images. A segmentation model is trained with the partially annotated samples to detect cancer regions, which are re-identified by a patch-wise classification network. Finally, the segmentation and classification results are combined to make the final image-wise classification. Several experiments are conducted on a landmark medical image dataset with partial annotations. We obtain a classification accuracy of 99.51%, which significantly outperforms other existing methods.

Keywords: Image classification · Pathological images · Convolutional neural network · Partial annotation

1 Introduction

Gastric cancer is one of the leading causes of cancer death worldwide [18]. Traditionally, pathologists must traverse through the entire pathological image to find lesions, but this process is time-consuming and fallible. Thus, computer-aided diagnostic systems are urgently required to reduce the burden of pathologists and improve the accuracy of diagnoses.

In recent years, deep learning has achieved remarkable success in the field of natural images [2,6,10,12], and it usually has a better performance than some traditional methods [7–9] in computer vision tasks. Considering the success of deep learning, many researchers, such as [1,4,5,16], attempted to apply it to the field of medical images. The performances of these applications largely rely

R. Hong et al. (Eds.): PCM 2018, LNCS 11165, pp. 476–486, 2018.
https://doi.org/10.1007/978-3-030-00767-6_44

on accurately annotated samples. However, it is often difficult to obtain a large amount of annotated samples in the field of medical images because of the high cost of accurate annotations.

To address these problems, a feasible solution is to make a compromise between data annotation and the performance. Thus, recent works proposed to use partially annotated samples in the medical image segmentation task [13]. The partially annotated samples only require the pathologists to annotate a part of the lesion areas in a medical image, which significantly reduces the cost of annotations and provides more detailed supervision information than the image-wise labels. However, in [13], they just iteratively generated new positive samples with high confidence to fine-tune the segmentation model. They ignored the rich information that spurious samples can provide which hardly affects the segmentation but greatly affects the classification. Therefore, we propose a new cascaded framework to classify partially annotated gastric pathological images in this paper. A segmentation model is trained with the partially annotated samples to detect cancer regions, which will be re-identified by a patch-wise classification network. Finally, the results of the segmentation network and classification network are combined to make the final image-wise classification. The major contributions of our work are as follows:

(1) We propose a new cascaded convolutional neural network framework to address the medical image classification problem with partial annotations; this framework is more efficient and practical than those of previous works;
(2) We propose a new training strategy to train the segmentation model by alternately using two complementary methods to sample the patches and considerably improve the classification performance;
(3) We further develop a medical image classification problem based on the segmentation problem by selecting patches that are difficult to distinguish by a segmentation model; then, we train the patch-wise classification model with these fallible patches. Thus, we combine the results of the segmentation model with the supervision information of the partial annotations to improve the performance in the image-wise classification task.

The remainder of our paper is organized as follows. Section 2 provides an introduction to closely related works. We provide the details of our proposed method in Sect. 3. Various experiments are conducted on a benchmark medical image dataset with partial annotations, and the results are reported in Sect. 4 which demonstrate the effectiveness of our method. In Sect. 5, we conclude our work and discuss the future works.

2 Related Work

Currently, the approaches of medical image classification can be divided into two main categories according to the datasets that they use: the approaches of image-wise labeled datasets and the approaches of pixel-wise labeled datasets. In the first category, the approaches often cannot obtain a good model because

of limited supervision information [14,15]. In the second category, they first generated patches according to the pixel-wise labels, which are used to train a classifier [17]. The classifier is further modified to handle the input of entire images and provide an image-wise classification result. Nevertheless, this method requires sufficient supervision information of pixel-wise annotations, which is of high expense. To address these problems, partially annotated data have also been used in medical image segmentation, where new positive samples are iteratively generated with high confidence to fine-tune the segmentation network, which won the 2017 China Big Data & Artificial Intelligence Innovation and Entrepreneurship Competitions[1]. These new positive samples can provide rich information, which are very helpful for segmentation. However, some spurious samples will be inevitably introduced in the process of generating new positive samples, which hardly affects the segmentation but greatly affects the classification. Therefore, we propose to integrate the information of partial annotations into the generating process of patches. Moreover, we propose to make use of the spurious samples in training our classification models. In the following section, we present a detailed introduction to the proposed approach.

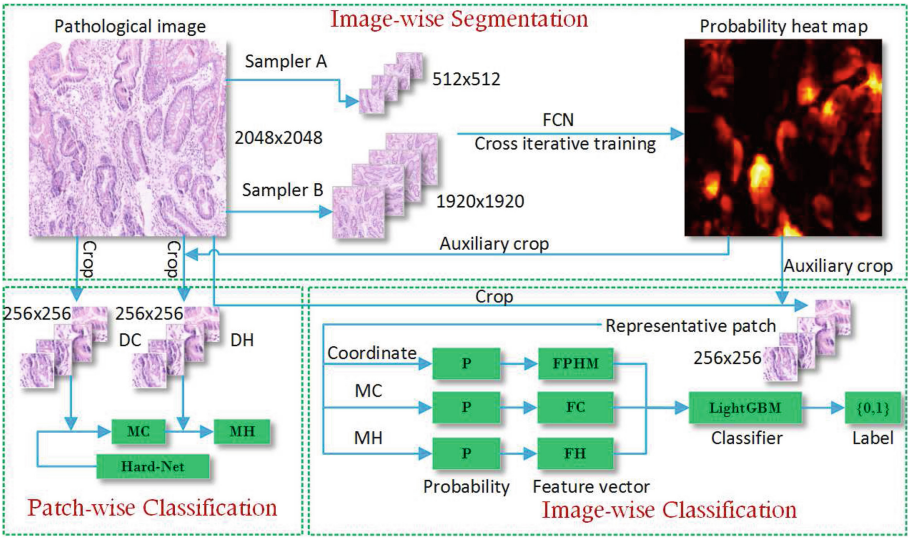


Fig. 1. Framework of the proposed approach. First, a segmentation model is trained with the partially annotated samples to detect cancer regions, which will be re-identified by a patch-wise classification network. Finally, we generate features from the segmentation and classification results and train a classifier to make the final decision.

[1] The challenge is held by Shanghai Big Data Alliance and Center for Applied Information Communication Technology (CAICT), the home page is http://www.datadreams.org/racerace3.html.

3 Proposed Approach

In this section, we describe the proposed method in details in three steps. The overall framework of our method is illustrated in Fig. 1. First, a segmentation network is trained with the partially annotated samples to obtain a heat map of the entire image, which provides the foundation for subsequent operations. We propose a new training strategy, which simultaneously considers two complementary sampling methods. Second, we develop a medical image classification problem based on the segmentation problem by generating the fallible patches from the segmentation heat map to train the patch-wise classification network. Finally, we present the image-wise classification process, which includes extracting features and training a classifier to obtain the final image-wise classification.

3.1 Image-Wise Segmentation

The main objective of the segmentation network is to make a pre-estimation of a given image. We use the existing segmentation network and propose a new training strategy to improve its performance. There are many classic segmentation networks such as [2,12,16]. In this paper, we use the FCN based on the pre-trained VGG16 model [19] as our segmentation network (**Seg-Net**) because pre-trained models can accelerate the convergence process and do not have stringent requirements for the segmentation details.

The cancer area of the original image is partially annotated, and the unannotated area is considered as a negative sample by default, which results in a large quantity of false-negative samples. We can address these obstacles by controlling the proportion of positive and negative samples. A smaller proportion of negative samples corresponds to a smaller risk of introducing false-negative samples. Meanwhile, adequate negative samples are required to train the Seg-Net. Considering these two factors, we propose two types of sampling methods.

For the first method (Sampler A), we sample the patches from the original images with the area ratio greater than 0.5. The area ratio P is defined as follows:

$$P = \frac{R}{R_{max}} \tag{1}$$

where R is the annotated cancer area in one patch, and R_{max} is the maximum among all R that we can obtain from the current image. Here, we use the area ratio to filter the patches instead of a fixed proportion of positive and negative samples as used in the previous work [13]. The reason is that the patches are unlikely sampled from those images with a small annotated cancer area with a fixed proportion of positive and negative samples, which are notably important to maintain the diversity of the samples. Even so, some positive samples of the annotated cancer area remain unused to train the segmentation network, which degrades the performance.

To address this problem, we propose the second sampling method (Sampler B) as a complement to the first method. Given an image from the partially

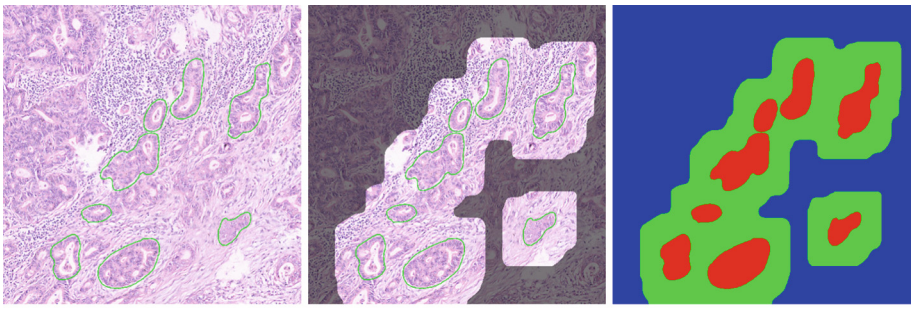


Fig. 2. An example generated by Sampler B. In the original image (a), the area circled by the green line is the cancer region, but there are some cancer regions in the remainder of the image. The highlighted area in image (b) is the valid area that contributes to the loss of the segmentation network. In image (c), the pixels in the red region are positive samples, and the pixels in the green region are negative samples. (Color figure online)

annotated dataset, the annotated cancer area is circled by the green line in the original image (Fig. 2(a)), which corresponds to the red area in Fig. 2(c); we denote this annotated area as R_{red}. However, some cancer areas remain unlabeled in the remainder of the image. The unannotated area near the annotated cancer area is less likely to contain false-negative samples. Considering these factors, we define the unannotated area near the annotated cancer areas as the security zone as follows:

$$|p_{sec} - p_{red}| < k \tag{2}$$

where p_{red} are the pixels in R_{red}, p_{sec} are the pixels in the security zone, and k is a parameter to control the proportion of negative samples. The pixels in the security zone are considered negative samples, and k is determined to satisfy the following:

$$\frac{1}{3} < \frac{R_{red}}{R_{green} + R_{red}} < \frac{2}{3} \tag{3}$$

where R_{green} is the area of the security zone (green region in Fig. 2(c)). The original image is further used as the input to train the segmentation network. We dismiss the pixels in R_{blue} when calculating the loss. We use a maximum filter with a kernel size $k \times k$ to generate an approximate security zone in implementation.

We propose a cross iterative training strategy that alternately uses two types of sampling methods in the training process. Thus, we make full use of the annotation information in the training process and can obtain a better performance.

3.2 Patch-Wise Classification

In this section, we introduce how to further develop a medical image classification problem based on the segmentation problem. First, we generate patches

from the original images to train the patch-wise classification network (**Hard-Net**). Then, we use the pre-estimation from Seg-Net to generate fallible patches to fine-tune the Hard-Net. Hence, we make an amendment about the samples indistinguishable by the Seg-Net.

First, we introduce the method to train the Hard-Net. Since the image is partially labeled, we only sample 256×256 patches whose centroids are located in the annotated cancer areas of image as positive samples. For negative samples, we randomly sample patches from non-cancer images. We do not sample patches from the unannotated areas of the cancer images because some cancer regions may remain unlabeled in these areas, which results in a large number of false-negative samples. Meanwhile, pathologists only label typical cancer nests, which results in potential false-positive samples. To address this problem, those patches are excluded, where the number of cells is lower than a threshold N_{cell} because more cells in the patch corresponds to a higher confidence that the patch is a positive sample. However, we cannot accurately calculate the number of cells in a patch. We approximate the number of cells using the area of cells. Since the cell densities vary with different images, it is difficult to find a fixed N_{cell} for all images. To overcome this obstacle, we simultaneously sample multiple patches from one image and only select the patch with the largest area of cells as the positive sample. Finally, we denote those patches as (**DC**). We select resnet50 [6] as our base model and obtain the Hard-Net by modifying the dimension of the output layer from 1000 to 2. We pre-train the Hard-Net on the DC and denote this model as **MC**. Then, we generate the patches that are indistinguishable by the segmentation network to fine-tune MC. Generally, the maximum value P_{max} of the probability heat map from the Seg-Net can represent the probability of cancer. However, we find a strange phenomenon that P_{max} is relatively small for some cancer images but relatively large for some non-cancer images, which causes difficulty in the final classification and urgently requires re-identification.

We first select non-cancer images with P_{max} greater than 0.20 and cancer images with P_{max} less than 0.70. Then, we sample the patches from these selected images. We sample the patches from non-cancer images according to the probability heat map in descending order as negative samples. Specifically, we sample the patches from the annotated cancer area according to the probability heat map in an ascending order as positive samples. However, it does not work well. We analyze this phenomenon and attribute it to the impurity of the images; the positive samples from the annotated cancer area with a relatively low P_{max} are likely the impurity part of the cancer nest. Therefore, we also sample positive samples based on the probability heat map in descending order. We denote those patches as **DH**, fine-tune the Hard-Net on DH and denote this model as **MH**.

3.3 Image-Wise Classification

In this section, we introduce the method to extract features from the high-confidence patches and the training details of the classifier to implement the final classification.

First, we sample N patches from the original images based on the probability heat map in descending order as representative patches with the constraint that the overlap between the patches cannot exceed 50%. Then, we obtain N probability values through MH and construct the following features:

$$f_m = \sum_{n=1}^{N} Sign\left(n, \frac{m}{M+1}\right) \quad 1 \leqslant m \leqslant M \tag{4}$$

$$Sign(n,t) = \begin{cases} 1 & p_n \geq t \\ 0 & p_n < t \end{cases} \tag{5}$$

where f_m is the m-th element in the feature vector, p_n is the n-th probability value, and we can obtain a feature vector with length M. M is set to 16 in our experiments. For a specific patch, we can obtain three probability values P_1, P_2 and P_3 from **Seg-Net**, **MC** and **MH**, respectively. More specifically, P_1 is a probability of the heat map, where we sample the representative patch. Therefore, we can obtain three feature vectors, **FPHM**, **FC** and **FH**, which correspond to three probability values respectively. Finally, we merge the three feature vectors into one feature vector with a length of 48, which is further applied to train the image-wise classifier by LightGBM [11].

4 Experiments

In this section, we conduct various experiments to verify the effectiveness of our proposed method. All experiments are implemented on the Keras framework [3]. We first briefly introduce the partially annotated dataset in the experiments. Then, we analyze the effect of the proposed training strategy, different feature combinations and number of patches. Finally, we compare our proposed method with previous works.

4.1 Gastric Tumor Dataset

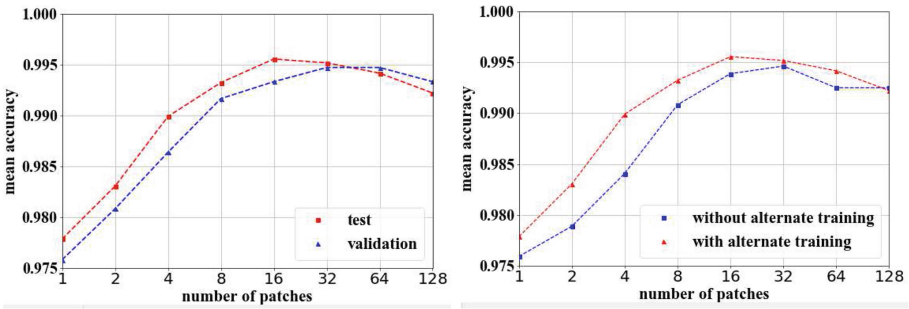
The Gastric Tumor Dataset was provided by 2017 China Big Data & Artificial Intelligence Innovation and Entrepreneurship, which contains 2100 images; 80% of the dataset is cancer images, whose cancer area is partially labeled. The images are available as 2048 × 2048 pixel TIFF images. We divide our dataset into training, validation and test sets with a split ratio of 3:1:3. To avoid randomness, we repeat the random splits for 4 times and report the average performance in our experiments. All parameters are selected on the validation set.

4.2 Effect of the Number of Representative Patches and Proposed Training Strategy

The number of representative patches that we use to extract features is a key parameter in our experiments. As shown in Fig. 3(a), a moderate accuracy is

Table 1. Comparison of different feature combinations in terms of mean accuracy ± standard deviation (%).

Methods	Accuracy	Dimension
FPHM	95.44 ± 0.56	16
FC	97.36 ± 0.39	16
FH	99.48 ± 0.22	16
FPHM + FC	97.41 ± 0.44	32
FPHM + FH	99.48 ± 0.21	32
FC + FH	99.50 ± 0.21	32
FPHM + FC + FH	$\mathbf{99.51} \pm 0.24$	48



(a) (b)

Fig. 3. The horizontal axis shows the number of representative patches sampled from one image, and the vertical axis shows the corresponding mean accuracy. In (a), the red (blue) line is the accuracy curve of the test (validation) data. In (b), the blue line is the accuracy curve without alternate training (only with Sampler A), and the red line stands for the accuracy curve with proposed alternate training (Samplers A and B). (Color figure online)

obtained with 32 representative patches; then, the accuracy decreases. The reason is that when there are too few patches, the collected information is insufficient, and when we select too many patches, the information is submerged by noise. A tradeoff must be made between the information and noise. From the results, we select 32 as the number of representative patches in our experiments according to the accuracy curve of the validation data.

We propose a cross iterative training strategy that alternately trains our segmentation model with two types of sampling methods. Fig. 3(b) shows us a comparison between different training strategies. From the result, we can easily conclude that the proposed training strategy improves the classification performance in most cases, which shows the effectiveness of the proposed training strategy.

4.3 Effect of Different Feature Combinations

This experiment analyzes the effect of different feature combinations. As mentioned, for a specific image, we can obtain three feature vectors: **FPHM**, **FC** and **FH**. Table 1 shows us the comparison between different combinations of these three feature vectors. We can conclude that FH outperforms the other two single features which is in consistent with our target. FH outperforms FPHM, which demonstrates that the re-identification of candidate cancer areas can help us improve the classification of the entire images. FH outperforms FC, which demonstrates the effectiveness of introducing fallible patches into training the patch-wise classification model. The fusion of features has a better performance, demonstrates that information from different perspectives can help us obtain a better performance of the final image-wise classification.

Table 2. Quantitative results of our proposed method and other recently published deep-learning-based methods in terms of mean accuracy ± standard deviation (%).

Methods	Accuracy	Precision	Recall	F1-score
WS [15]	95.22 ± 0.34	95.28 ± 0.46	98.70 ± 0.29	96.96 ± 0.25
End2End [17]	94.26 ± 1.02	95.69 ± 2.32	97.14 ± 1.34	94.38 ± 0.70
P-SEG [13]	97.89 ± 0.98	98.89 ± 0.83	98.49 ± 0.77	98.69 ± 0.61
Ours	$\mathbf{99.51} \pm 0.24$	$\mathbf{99.62} \pm 0.29$	$\mathbf{99.76} \pm 0.21$	$\mathbf{99.69} \pm 0.16$

4.4 Comparison with Existing Works

In this section, we compare our proposed method with three state-of-the-art methods (**WS** [15], **P-SEG** [13] and **End2End** [17]) to demonstrate the effectiveness of our method. The results are shown in Table 2. Our proposed method outperforms other methods in terms of mean accuracy, precision, recall, and F1-score. WS works with medical datasets with image-wise labels, and it does not make full use of the supervision information. End2End is applied to datasets with pixel-wise labels, and its performance largely relies on the quantity of pixel-wise labeled training samples. For the partially annotated datasets, they will suffer from the lack of labeled samples, which tremendously degrades the performance. P-SEG handles the segmentation problem of partially annotated datasets and iteratively generates new positive samples with high confidence to fine-tune the patch-wise classification model. However, some spurious samples may be introduced in the process of generating new positive samples. In our proposed method, we develop a medical image classification problem based on the segmentation problem by selecting patches that are indistinguishable by the segmentation model; then, we fine-tune the classification model with these fallible patches. Thus, we make full use of the supervision information of the partial annotations and obtain an outstanding performance with our method.

5 Conclusions

In this paper, we propose a cascaded convolutional neural network framework to classify partially annotated images. The proposed method achieves an outstanding performance on partially annotated medical image datasets. The proposed approach can prominently reduce the limitations of histopathology research, which will further promote the research of biomedical histopathology classification problems.

Acknowledgments. This work was supported by National Key Research and Development Program of China 2017YFB1002203, NSFC No. 61572451, and No. 61390514, Fok Ying Tung Education Foundation WF2100060004 and Youth Innovation Promotion Association CAS CX2100060016.

References

1. Chen, H., Qi, X., Yu, L., Heng, P.A.: DCAN: deep contour-aware networks for accurate gland segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2487–2496 (2016)
2. 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**(4), 834–848 (2018)
3. Chollet, F.: Keras. https://github.com/fchollet/keras (2015)
4. Cireşan, D.C., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Mitosis detection in breast cancer histology images with deep neural networks. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8150, pp. 411–418. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40763-5_51
5. Dhungel, N., Carneiro, G., Bradley, A.P.: The automated learning of deep features for breast mass classification from mammograms. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 106–114. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_13
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778 (2016)
7. Hong, R., Hu, Z., Wang, R., Wang, M., Tao, D.: Multi-view object retrieval via multi-scale topic models. IEEE Trans. Image Process. **25**(12), 5814–5827 (2016)
8. Hong, R., Zhang, L., Tao, D.: Unified photo enhancement by discovering aesthetic communities from flickr. IEEE Trans. Image Process. **25**(3), 1124–1135 (2016)
9. Hong, R., Zhang, L., Zhang, C., Zimmermann, R.: Flickr circles: aesthetic tendency discovery by multi-view regularized topic modeling. IEEE Trans. Multimed. **18**(8), 1555–1567 (2016)
10. Huang, G., Liu, Z., Weinberger, K.Q., van der Maaten, L.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, p. 3 (2017)
11. Ke, G., et al.: LightGBM: a highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems, pp. 3149–3157 (2017)

12. 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)
13. Nan, Y., et al.: Partial labeled gastric tumor segmentation via patch-based reiterative learning. arXiv preprint arXiv:1712.07488 (2017)
14. Paeng, K., Hwang, S., Park, S., Kim, M.: A unified framework for tumor proliferation score prediction in breast histopathology. In: Cardoso, M.J., et al. (eds.) DLMIA/ML-CDS -2017. LNCS, vol. 10553, pp. 231–239. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67558-9_27
15. Rakhlin, A., Shvets, A., Iglovikov, V., Kalinin, A.A.: Deep convolutional neural networks for breast cancer histology image analysis. arXiv preprint arXiv:1802.00752 (2018)
16. 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
17. Shen, L.: End-to-end training for whole image breast cancer diagnosis using an all convolutional design. In: Neural Information Processing Systems 2017 Workshop on Machine Learning for Health (2017)
18. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics. CA: Cancer J. Clin. **66**(1), 7–30 (2016)
19. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)