M-SEAM-NAM: Multi-instance Self-supervised Equivalent Attention Mechanism with Neighborhood Affinity Module for Double Weakly Supervised Segmentation of COVID-19

Wen Tang[1], Han Kang[1], Ying Cao[1], Pengxin Yu[1], Hu Han[2], Rongguo Zhang[1(✉)], and Kuan Chen[1]

[1] InferVision Medical Technology Co., Ltd., Beijing, China
zrongguo@infervision.com
[2] Institute of Computing Technology, Chinese Academy of Science, Beijing, China

Abstract. The Coronavirus Disease 2019 (COVID-19) pandemic has swept the whole world since 2019. Chest computed tomography (CT) plays an important role in clinical diagnosis, management and progression monitoring of COVID-19 patients. In order to decrease the cost of manual segmentation, weakly supervised segmentation methods, such as class activation maps (CAM) based methods, have been applied to achieve COVID-19-related lesion segmentation. Such methods could be used to localize the lesion preliminarily, but it is not precise enough to segment the lesion. In this paper, we propose a double weakly supervised segmentation method to achieve the segmentation of COVID-19 lesions on CT scans. A self-supervised equivalent attention mechanism with neighborhood affinity module is proposed for accurate segmentation. Multi-instance learning is adopted for training using annotations weaker than image-level. A simple pre-training process is also proved to be effective. We achieve a higher average Dice compared to Unet (0.782 vs 0.601) on COVID-19 lesion segmentation tasks. Codes in this paper will be available at https://github.com/TangWen920812/M-SEAM-NAM.

Keywords: Weakly supervised segmentation · Multi-instance learning · COVID-19

1 Introduction

Coronavirus Disease 2019 (COVID-19) has been announced as a global pandemic by the World Health Organization (WHO) [20]. By December 21th, over 75 million confirmed cases and 1 million deaths are reported across the globe [19]. Early detection, timely isolation and treatment of patients are advocated by WHO in order to control the disease transmission. Chest computed tomography (CT) plays an important role in the identification of suspected patients and

M. de Bruijne et al. (Eds.): MICCAI 2021, LNCS 12907, pp. 262–272, 2021.
https://doi.org/10.1007/978-3-030-87234-2_25

could provides quantitative evaluation of disease progression. It is listed in the diagnosis and treatment guidelines of the diagnosis and treatment of COVID-19 issued by many countries, such as China [14], Japan [8], and the United of Kingdom [3].

Deep learning methods are widely used to process medical image to assist the detection and segmentation of COVID-19. Several studies [11,17,23]have proven the effectiveness of Convolution Neural Network(CNN) to differentiate CT containing COVID-19 lesions or not. These models could help with timely patient triage but not quantitative analysis. COVID-19 related lesion segmentation could assist the localization of radiological abnormalities, and is the basis for further quantitative analysis of lesion area. Some Unet-based [16] methods [2,7,9,15] are applied to assist lesion segmentation for quantitative analysis. However, big amount of manual annotations are required for model training, which are time- and energy- consuming. Some studies adopt weakly supervised methods for COVID-19 lesion segmentation. Issam et al. [10] propose a weakly supervised method using detection or key point annotation to decrease the labeling cost. Yet the label used for weakly supervised segmentation is still not as simple as classification label. Hu et al. [6] use image-level labels and class activation maps (CAM) [24] to perform COVID-19 lesion segmentation. However, ignorance of multiple scales and details in CAM makes this method not robust enough among variable lesion sizes. Wang et al. [18] propose a new method focusing on variable sizes of targets, and achieve state-of-the-art performance with only image-level annotation on PASCAL VOC 2012 [5]. But there is no evidence shows [18] could work on COVID-19 dataset. Comparing to nature image dataset, COVID-19 data-set needs a larger number of more professional annotations. Multi-instance learning [13], a training method wherein a bag of images share one label, is a good solution requiring weaker annotations. It is usually used on huge image input which could not be put into model directly because of memory limitation. Thus, in this method, huge images are cut into several small image bags that share one label [1,22].

In this paper, we develop an end-to-end model, Multi-instance Self-supervised Equivalent Attention Mechanism with Neighborhood Affinity Module (M-SEAM-NAM), for doubly weakly supervised segmentation on COVID-19 dataset. We propose a new neighborhood affinity module on self-supervised equivalent attention mechanism to achieve better performance on lesions with variable sizes. We also adopt multi-instance learning in our model to use annotations weaker than image-level. Such designs allow us to train the model with doubly weakly classification labels while achieving better performance than fully supervised methods.

2 Method

We will introduce our method from the following three parts: the baseline method (Sect. 2.1), the improvement of our neighborhood affinity to the baseline (Sect. 2.2) and the combination of multi-instance learning with our proposed model (Sect. 2.3).

2.1 Self-supervised Equivalent Attention Mechanism (SEAM)

There is evidence [12] shows that different scales of an image would produce very different class activation maps, and thus offer different contextual information. This property has been leveraged in multi-scale CAM [12] to perform weakly supervised segmentation. However, it requires batch of network structures and complicated post-processing. Such issues are recently solved by self-supervised equivalent attention mechanism(SEAM) [18] which we will introduce in following paragraphs.

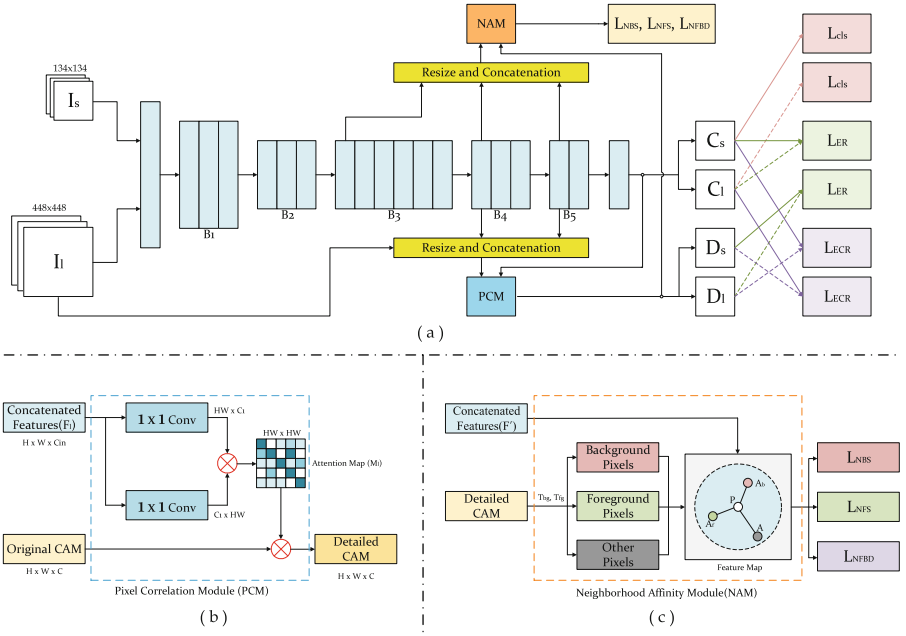

Fig. 1. (a) Overall structure of the proposed network. (b) Pixel Correlation Module (PCM). (c) Neighborhood Affinity Module (NAM).

Firstly, we define our input images and corresponding classification labels as $\{I^i\}_{i=1...n}$ and $\{y^i\}_{i=1...n}$. As shown in Fig. 1, each image I^i is sampled into two different scales: a large scale image I_l^i and a small scale image I_s^i. I_l^i and I_s^i are then put into the shared parameters backbone, ResNet38 [21], to gain two class activation maps, C_l^i and C_s^i. Then a Pixel Correlation Module (PCM) is used to produce detailed class activation maps using self-attention mechanism. I_l^i is downsampled to the size of feature maps on $B4$ and concatenated with feature maps on $B4$ and $B5$ in ResNet38 to form the concatenated feature maps: F_l^i. After that, as shown in Fig. 1 (b), F_l^i is compressed by a 1×1 convolution and flattened to calculate the non-local self-attention maps: M_l^i. Beacause F_l^i combines low-level and high-level information, M_l^i could give more reasonable

details. We can get more detailed class activation maps: D_l^i and D_s^i by multiplying C_l^i and C_s^i with M_l^i.

Three losses are used in SEAM. To assist the network to gain information from different scales, the equivariant regularization loss (L_{ER}) is used between C_l^i and C_s^i, as well as D_l^i and D_s^i which should be the same regardless of scale change. To ensure the PCM could work, the equivariant cross regularization loss (L_{ECR}) is used between C_l^i and D_s^i, as well as C_s^i and D_l^i. Using L_{ECR}, D_l^i and D_s^i could locate the basic activation area, and would not make the PCM fall into local minimum. In addition, we also use soft margin loss as the classification loss (L_{cls}) to supervise the whole network training.

$$L_{cls} = log(1 + e^{-y^i * p_l^i}) + log(1 + e^{-y^i * p_s^i}) \tag{1}$$

$$L_{ER} = \frac{1}{N_s^i} \parallel Down(C_l^i) - C_s^i \parallel_1 + \frac{1}{N_s^i} \parallel Down(D_l^i) - D_s^i \parallel_1 \tag{2}$$

$$L_{ECR} = \frac{1}{N_s^i} \parallel Down(C_l^i) - D_s^i \parallel_1 + \frac{1}{N_s^i} \parallel C_s^i - Down(D_l^i) \parallel_1 \tag{3}$$

where p_l^i and p_s^i are the global average pooling result of C_l^i and C_s^i, respectively. N_s^i is the number of pixels in C_s^i. $\parallel \bullet \parallel_1$ is the L1 distance and $Down(\bullet)$ is a downsampling operation. In addition, the SEAM is pre-trained on natural images using ImageNet [4]. We use a new pre-training method for SEAM on medical images. We use CAM method to pre-train ResNet38 backbone in SEAM. In detail, we remove PCM along with L_{ER} and L_{ECR} in SEAM and only train the backbone with L_{cls}. The experiment in Sect. 3.3 shows the effectiveness of our operations.

2.2 Neighborhood Affinity Module (NAM)

Although SEAM focuses on different scales of images, it is still not good enough to solve the COVID-19 segmentation problem, as the overall changes of natural images are bigger than medical images. Such difference allows the model trained on natural images to perceive more information than the model trained on medical images, which only focuses on one significant feature. As shown in Fig. 2, SEAM mainly focuses on the edge area of large lesions and produces false positive predictions around small lesions. Based on the observation, we believe that enhancing the relevance of features from neighborhood pixels would help to improve the model performance. So we introduce a neighborhood affinity module (NAM) to the basic SEAM.

Firstly, we define the position of pixels in D_l^i as $P = \{(x^j, y^j)\}_{j=1...J_i}$, and the position of pixels around (x^j, y^j) within a radius of $r(r = 5)$ as $A_j = \{x^k, y^k\}_{k=1...K_j}$. Then, the prediction result of one pixel is defined as $\sigma(D_l^i)[x^j, y^j]$, where $\sigma(\bullet)$ is the sigmoid activation function. Considering the influence caused by uncertain pixels on network optimization, two thresholds, T_{fg}, and T_{bg}, are defined on $\sigma(D_l^i)$ to categorize P into three groups: background pixels $P_b = \{(x^j, y^j) \mid \sigma(D_l^i)[x^j, y^j] < T_{bg}\}_{j=1...J_i} = \{P_b^j\}_{j=1...J_b^i}$,

foreground pixels $P_f = \{(x^j, y^j) \mid \sigma(D_l^i)[x^j, y^j] > T_{fg}\}_{j=1...J_i} = \{P_f^j\}_{j=1...J_f^i}$ and other uncertain pixels which would not be used. We can also get $A_b^j = \{(x^k, y^k) \mid \sigma(D_l^i)[x^k, y^k] < T_{bg} \cap (x^k, y^k) \in A^j\}_{k=1...K^j} = \{A_b^j(k)\}_{k=1...K_b^j}$ and $A_f^j = \{(x^k, y^k) \mid \sigma(D_l^i)[x^k, y^k] > T_{fg} \cap (x^k, y^k) \in A^j\}_{k=1...K^j} = \{A_f^j(k)\}_{k=1...K_f^j}$ by the same way. After that, a concatenation ($F^{'i}$) of three different features on $B3$, $B4$ and $B5$, as shown in Fig. 1, is put into NAM. Based on the definition of P_b, P_f, A_b^j, and A_f^j , we can sample the features of each kind of pixels as $F^{'i}[P_b^j]$, $F^{'i}[P_f^j]$, $F^{'i}[A_b^j(k)]$, and $F^{'i}[A_f^j(k)]$.

To enhance the relevance of features in neighborhood, we propose three loss, neighborhood foreground similarity loss (L_{NFS}) measuring similarity between foreground and foreground, neighborhood background similarity loss (L_{NBS}) measuring similarity between background and background, and neighborhood fore-back distinctive loss (L_{NFBD}) measuring the difference between foreground and background:

$$L_{NFS} = \frac{1}{J_f^i}\sum_j\left(\frac{1}{K_f^j}\sum_k(1 - cos(F^{'i}[P_f^j], F^{'i}[A_f^j(k)]))\right) \tag{4}$$

$$L_{NBS} = \frac{1}{J_b^i}\sum_j\left(\frac{1}{K_b^j}\sum_k(1 - cos(F^{'i}[P_b^j], F^{'i}[A_b^j(k)]))\right) \tag{5}$$

$$\begin{aligned} L_{NFBD} = &\frac{1}{J_f^i}\sum_j\left(\frac{1}{K_b^j}\sum_k(cos(F^{'i}[P_f^j], F^{'i}[A_b^j(k)]))\right) \\ &+\frac{1}{J_b^i}\sum_j\left(\frac{1}{K_f^j}\sum_k(cos(F^{'i}[P_b^j], F^{'i}[A_f^j(k)]))\right) \end{aligned} \tag{6}$$

where $cos(\bullet, \bullet)$ is the function of cosine similarity.

2.3 Multi-instance Network with SEAM and NAM

The SEAM with NAM still requires a large amount of classification annotations. We adopt the multi-instance training idea in our model to implement a doubly weakly segmentation network which could use weaker classification labels to achieve good segmentation performance. Radiologists only need to give a rough slice range of lesion when labeling. For example, radiologists note that there are lesions from 100 to 110 slices but do not need to record the exact slice numbers. In this way, weaker classification labeling costs less time and we could use some of the unknown layers in training.

The multi-instance method is usually used in pathological image. It helps to solve the problem that a lot of images share one label. However, if the multi-instance method could work, the positive data batch must contain at least one positive image. In our situation, the backbone of our network is a segmentation backbone, which means we can not use patient classification label because of memory limitation. Thus in our experiments, we define positive batch and negative batch based on three categories of slice described in Sect. 3.1. We select

one positive layer and other seven consecutive layers (could be unknown layers or positive layers) as one positive data batch. We also randomly select eight consecutive images from negative patients as a negative batch. Because each batch has only one label, we keep other losses all the same and change the classification loss to Eq. 7, where y is the label of one batch, p^i is the prediction of one input image, and n is batch size. In addition, we update the network every eight batches to avoid jumping changes on loss and also change the CAM pre-training loss as Eq. 7.

$$L_{cls} = -log(1 + e^{-y*pb}) \ , \ pb = log(\frac{1 - \prod_i^n (1 - \sigma(p^i)}{\prod_i^n (1 - \sigma(p^i)}) \tag{7}$$

$$L_{all} = L_{cls} + L_{ER} + L_{ECR} + L_{NBS} + L_{NFS} + 0.5 * L_{NFBD} \tag{8}$$

3 Experiments

3.1 Data Description

The COVID-19 CT dataset used in this study is collected from two hospitals. All positive cases are confirmed by RT-PCR and show lesions related to COVID-19 on CT confirmed by radiologists, while all negatives cases are also confirmed by RT-PCR and without lesions related to COVID-19 on CT. The lesions related to COVID-19 indicate the imaging features of COVID-19 pneumonia including multiple small patchy shadows, interstitial changes appear, multiple ground-glass shadows, infiltrates shadows, and pulmonary consolidation. There are 587 positive cases and 288 negative cases from the first hospital. These cases are further divided into a training set (522 positive and 240 negative cases), and a testing set (65 positive and 48 negative cases). Cases collected from the second hospital are used for testing only, including 68 positive and 49 negative cases. Lesions related to COVID-19 of all positive cases in testing set (65 + 68 patients) are annotated by two experienced radiologists on all layers, and lesions of positive cases in training set (522 patients) are annotated every four or five layers. In all, 8309 out of 198882 CT scans layers were labeled with ***Lesion*** in the training set and 4725 + 5303 out of 24304 + 25447 CT scans layers are labeled with ***Lesion*** in the testing set. These segmentation annotations in training set are used to train fully supervised models. M-SEAM-NAM is trained using the positive/negative classification labels.

The fully supervised model used in experiments is trained using CT layers labeled with segmentation annotations and the same amount of negative layers in negative patients. The classification labels of training set are classified into 3 categories: 1) positive layers, layers with segmentation annotations; 2) negative layers, layers from negative patients; and 3) unknown layers, layers from positive patients without any annotation. The weakly supervised models we used are trained using the positive/negative classification labels. Negative layers are randomly selected to keep the sample balance.

3.2 Overall Implementations

The proposed method is implemented using Pytorch. The losses of the network are optimized by SGD, which is a method for stochastic optimization. The learning rate is 0.001 with linear decline and the model is trained for 16 epochs. Other weakly supervised segmentation methods for comparison are also trained under the same setting. The fully supervised segmentation method (Unet) is trained for 100 epochs. Common data augmentations, including shift, rotation, flip, brightness changing and center cropping are utilized during training. All CT layers are resized to 448 × 448 before inputting to the network. Additionally, we use 3 consecutive layers images as input. All the networks are trained in 2D, whereas evaluation indexes are calculated at patient level. Dice score, lesion pixel recall and lesion pixel precision are used to evaluate the models.

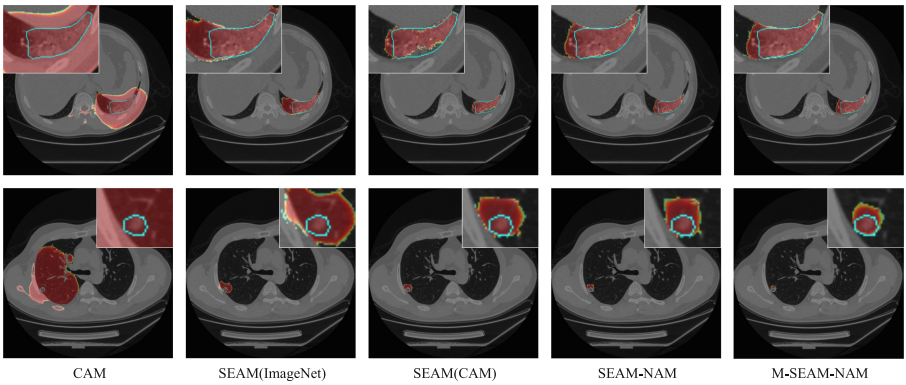

Fig. 2. Segmentation result on several weakly supervised methods. Red heatmaps are the prediction and blue contours are the ground truth. (Color figure online)

3.3 Model Comparison

We compare our method with several weakly-supervised methods including CAM, multi-scale CAM and SEAM(ImageNet pre-trained, baseline). To show the effectiveness of our proposed pre-training method, we use two different pre-trained parameters on SEAM. We also compare our model with Unet, a common fully supervised segmentation method. The dataset is split to show more detailed performance of each method, as positive cases could be used to show the model's segmentation performance and sensitivity, while negative cases could reflect the model's specificity. As shown in Table 1, SEAM pre-trained with CAM performs significantly better than SEAM pre-trained with ImageNet. Our SEAM with NAM model achieves a better result than the SEAM without NAM. It proves NAM is helpful on the segmentation of COVID-19. Our M-SEAM-NAM model outperforms all other models regarding positive, negative and all patients.

Because a big amount of unknown layers are included in the training process and this operation could increase the model's understanding of unlabeled layers, even with a weaker annotation, we still obtain the best result using the proposed model. As shown in Fig. 2, our pre-training process, NAM and multi-instance learning are all helpful for increasing true positive and decreasing false positive in lesions with variable sizes.

Table 1. Segmentation model comparison on COVID-19 testing dataset. In negative patients, if no lesions are segmented, the dice coefficient would be 1; otherwise, 0. Wilcoxon signed rank test is used to perform statistical tests. $T_{fg} = 0.1, r = 5$ is used on all method with NAM.

		Positive patients			Negative patients	All patients
Method	Pre-trained	Dice	Recall	Precision	Dice	Dice
Unet	-	0.600 ± 0.273	0.683 ± 0.280	0.574 ± 0.260	0.604 ± 0.489	0.601 ± 0.344 ($p < 0.001$)
CAM	-	0.109 ± 0.084	$\mathbf{0.974 \pm 0.04}$	0.060 ± 0.049	0.958 ± 0.200	0.336 ± 0.396 ($p < 0.001$)
Multi-CAM	-	0.509 ± 0.182	0.673 ± 0.103	0.470 ± 0.190	0.938 ± 0.242	0.637 ± 0.277 ($p < 0.001$)
SEAM (baseline)	ImageNet	0.495 ± 0.190	0.836 ± 0.127	0.383 ± 0.187	0.958 ± 0.200	0.619 ± 0.281 ($p < 0.001$)
SEAM	CAM	0.631 ± 0.139	0.834 ± 0.151	0.534 ± 0.160	0.958 ± 0.200	0.718 ± 0.214 ($p < 0.001$)
SEAM with NAM	CAM	0.683 ± 0.135	0.763 ± 0.121	0.634 ± 0.160	0.958 ± 0.200	0.757 ± 0.197 ($p < 0.001$)
multi-instance SEAM with NAM	CAM	$\mathbf{0.710 \pm 0.114}$	0.712 ± 0.124	$\mathbf{0.714 \pm 0.117}$	$\mathbf{0.979 \pm 0.143}$	$\mathbf{0.782 \pm 0.171}$

3.4 Ablation Experiment

As there are two thresholds in the neighborhood affinity module, we use ablation experiment to study their effect. Firstly, we analyse the statistical distribution of SEAM model output, i.e. the pixel probability value. Based on the statistical distribution (left histogram in Fig. 3), we set T_{bg} as 0.01 because it is confident that a pixel with prediction probability smaller than $T_{bg} = 0.01$ is background pixel. So that we only have to do ablation experiment on T_{fg}. As shown in the line chart in the right of Fig. 3, we set T_{fg} from 0.1 to 0.8. The reason we do not use 0.9 as one threshold is that when $T_{fg} = 0.9$, there is no foreground pixels at the beginning of the training process. According to the results, we choose to use $T_{fg} = 0.1$. Another hyperparameter is the neighborhood radius. As shown in Fig. 3, 5 is the best radius. According to the results, the values of two thresholds and radius have slight influence on model performance.

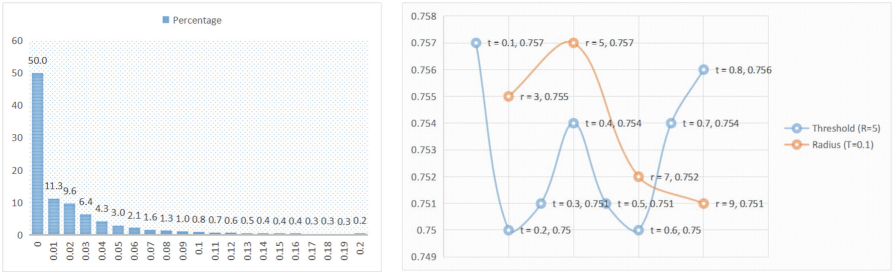

Fig. 3. Ablation experiment result

4 Conclusion

We introduce a SEAM method for COVID-19 weakly supervised segmentation and propose to use CAM model as a pre-trained model which performs better than models pre-trained with ImageNet. We also propose a NAM method to solve the problem that SEAM model performs unsatisfactorily on segmentation of lesion with variable sizes in COVID-19 datasets. The SEAM-CAM method we proposed performs best among all models. Considering the time and labor cost of annotation, we also conduct a simple labeling strategy that radiologists just label the approximate slice range of lesions other than the exact class of every single slice. By using this weaker classification labeled data, we train a doubly weakly supervised segmentation model, M-SEAM-CAM, via multi-instance learning. Our proposed method achieves an better performance because of the effective use of unlabeled data. In future study, we will try to compress our model and use patient-level classification annotation to have a much weaker supervised model.

References

1. Cances, L., Pellegrini, T., Guyot, P.: Sound event detection from weak annotations: weighted-gru versus multi-instance-learning. In: Proceedings of the Detection and Classification of Acoustic Scenes and Events 2018 Workshop (DCASE2018), pp. 64–68 (2018)
2. Cao, Y., et al.: Longitudinal assessment of covid-19 using a deep learning-based quantitative ct pipeline: illustration of two cases. Radiol. Cardiothorac. Imaging **2**(2), e200082 (2020)
3. Chua, F., et al.: The role of ct in case ascertainment and management of covid-19 pneumonia in the uk: insights from high-incidence regions. Lancet Respir. Med. **8**(5), 438–440 (2020)
4. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009)

5. Everingham, M., Winn, J.: The pascal visual object classes challenge 2012 (voc2012) development kit. Pattern Analysis, Statistical Modelling and Computational Learning, Technical report 8 (2011)
6. Hu, S., et al.: Weakly supervised deep learning for covid-19 infection detection and classification from ct images. IEEE Access **8**, 118869–118883 (2020)
7. Huang, L., et al.: Serial quantitative chest ct assessment of covid-19: a deep learning approach. Radiol. Cardiothorac. Imaging **2**(2), e200075 (2020)
8. Japanese society for infection prevention and control. guide for responding to new coronavirus infections at medical institutions (ver.2.1). http://www.kankyokansen.org/uploads/uploads/files/jsipc/COVID-19_taioguide2.1.pdf. Accessed 5 Apr 2020
9. Jin, S., et al.: Ai-assisted ct imaging analysis for covid-19 screening: Building and deploying a medical ai system in four weeks. MedRxiv (2020)
10. Laradji, I., et al.: A weakly supervised region-based active learning method for covid-19 segmentation in ct images. arXiv preprint arXiv:2007.07012 (2020)
11. Li, L., et al.: Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct. Radiology (2020)
12. Ma, X., Ji, Z., Niu, S., Leng, T., Rubin, D.L., Chen, Q.: Ms-cam: multi-scale class activation maps for weakly-supervised segmentation of geographic atrophy lesions in sd-oct images. IEEE J. Biomed. Health Inform. **24**(12), 3443–3455 (2020)
13. Maron, O., Lozano-Pérez, T.: A framework for multiple-instance learning. Advances in Neural Information Processing Systems, pp. 570–576 (1998)
14. National health commission of the people's republic of china. chinese clinical guidance for covid-19 pneumonia diagnosis and treatment (7th edition). http://www.nhc.gov.cn/yzygj/s7653p/202003/46c9294a7dfe4cef80dc7f5912eb1989/files/ce3e6945832a438eaae415350a8ce964.pdf. Accessed 5 Apr 2020
15. Qi, X., et al.: Machine learning-based ct radiomics model for predicting hospital stay in patients with pneumonia associated with sars-cov-2 infection: A multicenter study. Medrxiv (2020)
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. Song, Y., et al.: Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images. MedRxiv (2020)
18. Wang, Y., Zhang, J., Kan, M., Shan, S., Chen, X.: Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12275–12284 (2020)
19. WHO: Coronavirus disease 2019 (covid-19) situation report. https://www.who.int/docs/default-source/coronaviruse/wou_21-dec_cleared.pdf?sfvrsn=a7575c1f_1&download=false. Accessed 21 Dec 2020
20. WHO: Who director-general's opening remarks at the media briefing on covid-19 - 11 March 2020. www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19--11-march-2020. Accessed 21 Dec 2020
21. Wu, Z., Shen, C., Van Den Hengel, A.: Wider or deeper: Revisiting the resnet model for visual recognition. Pattern Recogn. **90**, 119–133 (2019)
22. Yao, J., Zhu, X., Huang, J.: Deep multi-instance learning for survival prediction from whole slide images. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 496–504. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_55

23. Zheng, C., et al.: Deep learning-based detection for covid-19 from chest ct using weak label. MedRxiv (2020)
24. 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)