Cross Domain Pulmonary Nodule Detection Without Source Data

Rui Xu[1], Yong Luo[1(✉)], and Yan Xu[2]

[1] Wuhan University, Wuhan 430072, China
{rui.xu,luoyong}@whu.edu.cn
[2] The University of Chicago, Chicago, IL 60637, USA
yanx1@uchicago.edu

Abstract. The model performance on cross-domain pulmonary nodule detection usually degrades because of the significant shift in data distributions and the scarcity of annotated medical data in the test scenarios. Current approaches to cross-domain object detection assume that training data from the source domain are freely available; however, such an assumption is implausible in the medical field, as the data are confidential and cannot be shared due to privacy concerns. Thus, this paper introduces source data-free cross-domain pulmonary nodule detection. In this setting, only a pre-trained model from the source domain and a few annotated samples from the target domain are available. We introduce a novel method to tackle this issue, adapting the feature extraction module for the target domain through minimizing the proposed General Entropy (GE). Specifically, we optimize the batch normalization (BN) layers of the model by GE minimization. Thus, the dataset-level statistics of the target domain are utilized for optimization and inference. Furthermore, we tune the detection head of the model using annotated target samples to mitigate the rater difference and improve the accuracy. Extensive experiments on three different pulmonary nodule datasets show the efficacy of our method for source data-absent cross-domain pulmonary nodule detection.

Keywords: Pulmonary Nodule Detection · Domain Adaptation · Source Free · Entropy Minimization · Model Reuse

1 Introduction

There has been much progress in various object detection tasks [13,15,16,28,37] with the prosperity of deep learning. In the medical field, detection algorithms are able to obtain performance comparable to that of clinical experts, e.g. pulmonary nodule detection [18,27,33,34], etc. Nonetheless, most of the approaches are based on the assumption that the training/source and test/target data come from similar distributions. This assumption restricts the application of these approaches in the real world, because there often exists nontrivial domain difference between the training data and the real-world test data; the domain shift causes significant performance degradation of the algorithms in the test/target

domain. Hence, a great deal of effort has been directed towards cross-domain object detection [1,2,5,7,12,23,32,38,39] in recent years to enhance the performance of the source model on the target domain.

However, current approaches for cross-domain object detection still contain an improper assumption for medical applications. They assume that the training samples from the source domain are freely accessible, while in reality, medical data are usually not shareable due to privacy issues and merely a pre-trained source model is accessible. What's more, acquiring and annotating medical data are both time-consuming and costly, resulting in limited training samples of the target domain, making cross-domain object detection in the medical field very challenging. Considering these two aspects, we present a realistic but demanding setting, source data-free cross-domain detection of lung nodule. In this scenario, merely a pre-trained source model and a few annotated samples from the target domain are available. As far as we know, this is the first work that tackles source data-absent cross-domain adaptation in the pulmonary nodule detection task.

The batch normalization (BN) [9] layers of a model normalize and modulate the features, and thus are closely tied to the model performance when there is a shift in data distribution. In cross-domain image classification and semantic segmentation tasks, some studies simply substitute the source batch statistics with the statistics of the current batch of the target domain [14]. Some studies combine the statistics of both source and target [36]. Some other studies [31,35] pay attention to the target statistics, and minimize entropy loss to optimize the affine parameters as well. Nevertheless, these methods are either too weak or not applicable for cross-domain object detection.

In our cross-domain pulmonary nodule detection setting, which does not rely on source data, we propose adapting to the target domain by reducing the entropy of the model predictions. However, the original entropy [25] only supports image classification and segmentation currently. We successfully solve this problem by extending entropy to its detection variant, termed General Entropy (GE). We choose entropy for its ability to quantify uncertainty and shifts, as low entropy predictions are all-in-all more reliable and high entropy predictions represent larger shifts. To better utilize the source information and efficiently adapt, we only optimize the affine parameters and estimate the target dataset-level statistics in the batch normalization layers via entropy minimization. This step enables us to learn a target-specific feature encoding module under the same detection head, without requiring access to the source data or the labels of the target data.

To enhance the detection performance further and alleviate the common problem of rater disagreement in the medical field, we also fine-tune the detection head of the model using annotated samples from the target domain.

Our primary contributions are summarized as follows:

– We establish a source data-free setting for cross-domain lung nodule detection, utilizing merely a well-trained source model and a limited number of labeled target samples.
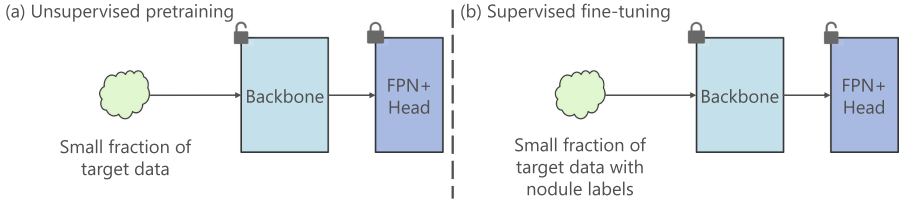
(a) Unsupervised pretraining

Small fraction of
target data

Backbone

FPN+
Head

(b) Supervised fine-tuning

Small fraction of
target data with
nodule labels

Backbone

FPN+
Head

Fig. 1. The pipeline of our proposed method. The source model is composed of a feature encoding module and a detection head module. (a) We keep the detection head frozen, and adapt the batch normalization (BN) layers in the feature extraction module by minimizing our Generalized Entropy (GE) to obtain target dataset-level statistics. (b) The detection head is fine-tuned using a small fraction of target data with labels.

- We propose a novel method, which adapts the model feature extraction module for the target domain via General Entropy (GE) minimization. We further fine-tune the model detection head with labeled target samples to improve the adaptation performance.
- For the purpose of evaluation, we curate a benchmark using four widely used pulmonary nodule datasets.

Experiments on the benchmark show our method can achieve the state-of-the-art results, demonstrating the effectiveness of our method.

2   Method

For a vanilla cross-domain adaptation (DA) task, we have N^s labeled samples $\{x_i^s, y_i^s\}_{i=1}^{N^s}$ from the source domain and also N^t labeled samples $\{x_i^t, y_i^t\}_{i=1}^{N^t}$ from the target domain. The main goal of DA is to address the domain shift between the source domain and the target domain, thus to well predict labels $\{y_i^t\}_{i=1}^{N^t}$ in the target domain. In this work, we assume that we cannot obtain samples from the source domain because of concerns related to privacy. Instead of the source dataset, we are given a well-trained source model $f_\theta(x)$ with parameters θ. Based on this assumption, we present source data-free cross-domain pulmonary nodule detection, and aim to learn a target model with the given well-trained source model $f_\theta(x)$ and target samples $\{x_i^t, y_i^t\}_{i=1}^{N^t}$.

Our method comprises two steps as shown in Fig. 1. First, the feature extraction module of the well-trained source model is adjusted to the target domain using unsupervised learning. To be specific, the batch normalization (BN) layers of the model are optimized by minimizing entropy loss to obtain target dataset-level statistics, where a general form of entropy termed Generalized Entropy (GE) is proposed. Then, using the annotated target samples, we further employ supervised learning to fine-tune the detection head of the model for rater difference mitigation and performance enhancement. In the following, we would like first to revisit two types of the uncertainty of the bounding box, the probability distribution representation and localization quality estimation, and then elaborate on our method in detail.

Preliminaries. There are two conventional representations for the bounding box $\mathcal{B}$ in detection. For instance, the central point coordinates, width, height, and depth, $\{a, b, c, w, h, d\}$ [3,17,21], and the distance from the sampling point to the up, down, top, bottom, left, and right planes, $\{u, d, t, b, l, r\}$ [28] are utilized to denote bounding boxes in the pulmonary nodule detection task. According to [37], there is no performance difference between the two representations. In this work, relative offsets from the sampling point to the six planes of a bounding box $\mathcal{B} = \{u, d, t, b, l, r\}$ are used as the regression targets, since the physical meaning of each variable in $\{u, d, t, b, l, r\}$ is consistent. Given the $\{a, b, c, w, h, d\}$ form, we will convert it to the $\{u, d, t, b, l, r\}$ form.

Yet this form follows the Dirac delta distribution that only concentrates on the ground-truth locations, and is too rigid to reflect the ambiguity of bounding boxes [6,13]. Recently, some works [13,20] adopt the probability distribution representation of the bounding box to learn its localization uncertainty. Let $y \in \mathcal{B}$ be the distance to a certain plane of a bounding box, whose estimated value $\hat{y}$ can be represented as:

$$\hat{y} = \int_{y_{min}}^{y_{max}} s \Pr(s) ds, \tag{1}$$

where s is the regression distance in range of $[y_{min}, y_{max}]$, and $\Pr(s)$ is the corresponding probability. Then, to be congenial with the convolutional neural networks, the continuous regression range $[y_{min}, y_{max}]$ is converted into a uniform discretized representation, $\{y_0, y_1, ..., y_i, y_{i+1}, ..., y_{n-1}, y_n\}$ with even intervals Δ, where $\Delta = y_{i+1} - y_i, \forall i \in [0, n-1]$, $y_0 = y_{min}$, and $y_n = y_{max}$. Thus, the estimated value $\hat{y}$ becomes:

$$\hat{y} = \sum_{i=0}^{n} \Pr(y_i) y_i, \tag{2}$$

where $\sum_{i=0}^{n} \Pr(y_i) = 1$, and the $\Pr(s)$ can be easily implemented using a SoftMax function with $n+1$ outputs. Hereto, the uncertainty of the bounding box offsets are modeled.

There is also another simple way to model the localization uncertainty of the bounding box, i.e. the localization quality estimation in the form of IoU [30] or centerness [28] score. Thereinto, the centerness [28] represents the distance measurement between the center points of the location and its corresponding object. Given the regression targets u^*, d^*, t^*, b^*, l^*, and r^* for a sampling point, the centerness $\hat{y}$ can be defined as:

$$\hat{y} = \sqrt{\frac{\min(u^*, d^*)}{\max(u^*, d^*)} \times \frac{\min(t^*, b^*)}{\max(t^*, b^*)} \times \frac{\min(l^*, r^*)}{\max(l^*, r^*)}}. \tag{3}$$

In our method, we employ the centerness [28] score measurement for its simplicity and good performance in pulmonary nodule detection.

2.1 Feature Extractor Adaptation

Entropy Objective. Our training goal is to reduce the entropy $H(\hat{y})$ of the model detection results $\hat{y} = f_\theta(x^t)$. This is because entropy is an unsupervised objective for uncertainty measurement, while related to the supervised task and model. However, the current Shannon entropy [25] only supports classification. Therefore, we propose Generalized Entropy (GE) that generalizes the Shannon entropy [25] for dense detectors. Assume that a model's final prediction $\hat{y}$ is the linear combination of two variables $\hat{y} = y_l p_{y_l} + y_r p_{y_r}, (y_l \leq \hat{y} \leq y_r)$, where $p_{y_l}, p_{y_r} (p_{y_l} \geq 0, p_{y_r} \geq 0, p_{y_l} + p_{y_r} = 1)$ are probabilities for these variables estimated by the model respectively. The proposed GE is able to cover the three special cases of the General Focal Loss (GFL) [13] for dense detectors:

When $\beta = \gamma, y_l = 0, y_r = 1, p_{y_r} = p, p_{y_l} = 1 - p$ and $y \in \{1, 0\}$ in GFL [13], GE for focal loss (FL) can be written as:

$$H(p) = -((1 - \alpha)p^\gamma(1 - p)\log(1 - p) + \alpha(1 - p)^\gamma p\log(p)). \tag{4}$$

When $y_l = 0, y_r = 1, p_{y_r} = \sigma$ and $p_{y_l} = 1 - \sigma$ in GFL [13], GE for quality focal loss (QFL) can be written as:

$$H(\sigma) = -(\sigma^\beta(1 - \sigma)\log(1 - \sigma) + (1 - \sigma)^\beta\sigma\log(\sigma)). \tag{5}$$

When $\beta = 0, y_l = y_i, y_r = y_{i+1}, p_{y_l} = \Pr(y_l) = \Pr(y_i) = \mathcal{S}_i$ and $p_{y_r} = \Pr(y_r) = \Pr(y_{i+1}) = \mathcal{S}_{i+1}$ in GFL [13], GE for distribution focal loss (DFL) can be written as:

$$H(\mathcal{S}_i, \mathcal{S}_{i+1}) = -(\mathcal{S}_i\log(\mathcal{S}_i) + \mathcal{S}_{i+1}\log(\mathcal{S}_{i+1})). \tag{6}$$

Modulation Parameters. As shown in Fig. 1, the pulmonary nodule detection network $f_\theta(x)$ is composed of two modules: the feature encoding module $g_\theta : x \to \mathbb{R}^d$ and the detection head module $h_\theta : \mathbb{R}^d \to \mathbb{R}^K$; $f_\theta(x) = h_\theta(g_\theta(x))$, d and K are dimensions of the extracted feature and the model output respectively. To keep the same hypothesis h_θ, a natural choice of the modulation parameters is all the feature extractor parameters g_θ; however, altering g_θ may cause the model to diverge from its training, since θ is the only representation of the source data in our setting. Besides, the limited number of training samples from the target domain is not suitable for optimizing the high dimensional θ. Previous works [31,35] find that adapting the batch statistics, especially dataset-level statistics, is effective for domain adaptation. Considering the feature modulation ability and low dimensional computation of the batch normalization (BN) layers, we choose to update the BN layers during training. Inside the BN layer, there are two sets of parameters: the statistics (μ, σ), which normalize the feature, and the affine parameters (β, γ), which modulate the feature. Given a batch of target samples $\{x_i^t\}_{i=1}^B$, where B is the batch size, the outputs of the BN layer $\{x_i^{t'}\}_{i=1}^B$ are calculated as:

$$x_i^{t'} = \gamma\overline{x_i^t} + \beta = \gamma\frac{x_i^t - \mu}{\sigma} + \beta,$$
$$\mu = \mathbb{E}[x_i^t], \sigma^2 = \mathbb{E}[(x_i^t - \mu)^2].$$

In the meantime, a running mean vector μ_r and a running variance vector σ_r are estimated using moving average to derive dataset-level statistics for the target domain:

$$\mu_r = \lambda\mu + (1-\lambda)\mu_r, \sigma_r^2 = \lambda\sigma^2 + (1-\lambda)\sigma_r^2. \tag{7}$$

The affine parameters (β, γ) are optimized via minimizing the GE loss.

2.2   Detection Head Adaptation

Transfer learning by fine-tuning is a common way to adjust a well-trained network to a new domain. To enhance the performance of pulmonary nodule detection even further, we tune the detection head of the model h_θ using the training samples from the target domain $\{x_i^t, y_i^t\}_{i=1}^{N^t}$. Meanwhile, this can also alleviate the issue of rater disagreement between different datasets, a common problem in the medical field.

3   Experiments

3.1   Benchmark and Evaluation

We establish a benchmark from PN9 [18] to LUNA16 [24]/tianchi [29]/russia [19] for shifts, as shown in Fig. 2. The specifics of these datasets are listed in Table 1. As seen, the CT scans in these datasets, which are gathered from various sites, have different image sizes and voxel sizes. In Table 2, we display the lung nodule size and quantity distribution of the four datasets.

Recall that vanilla domain adaptation requires the use of the labeled source data, while our setting denies the use of source data PN9 [18] during adaptation. We take into account only those CT scans having publicly available nodule annotations. The annotation files of the four datasets are csv files. Each line of the files holds the information of one nodule, including the CT scan filename it belongs to, and its location. In the three target datasets, the nodule location is indicated by the center coordinates and diameter, whereas in PN9 [18], it is marked by the top-left and bottom-right coordinates.

LUNA16 [24], tianchi [29], and russia [19] are divided into 7/1/2 for training, validation, and testing. In these three datasets, the raw CT data undergoes three pre-processing steps: 1) We use lungmask [8] to extract lung regions from each CT image and mask other regions to minimize irrelevant calculations. In this process, the HU values of the raw CT data are clipped into the range $[-1200, 600]$ and then linearly converted into the range $[0, 255]$, resulting in uint8 values. Then we set a padding value of 170 for regions outside the lung masks. 2) To prevent an excess of unnecessary hyper-parameters, the spacing of all the CT images is resampled to $(1.00, 1.00, 1.00)$ mm, ensuring consistency for the anchor design across all detectors. 3) To further improve the computational efficiency, we crop
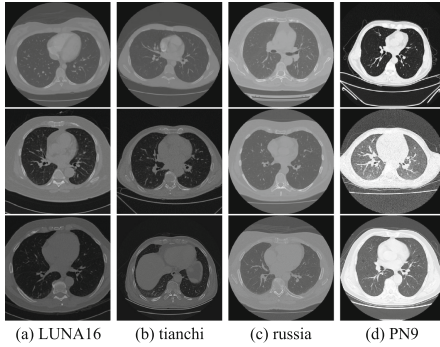
(a) LUNA16   (b) tianchi   (c) russia   (d) PN9

Fig. 2. Samples from four lung nodule datasets are shown, with each column corresponding to a dataset as marked. CT images from different datasets exhibit domain discrepancy, for instance, color contrast/saturation, voxel intensity, image spacing, amount of nodules.
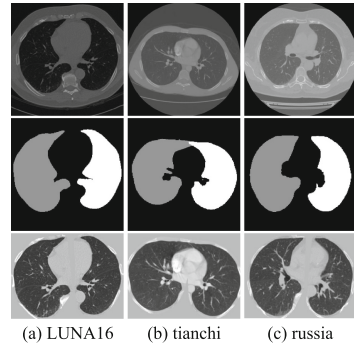

(a) LUNA16   (b) tianchi   (c) russia

Fig. 3. Samples of the pre-processed images in the LUNA16, tianchi, and russia. The 1st row contains the raw images, the 2nd row shows the extracted lung regions, and the 3rd row displays the pre-processed images.

Table 1. Pulmonary nodule datasets. 'Scans' and 'Class' indicates the number of CT scans and the class, respectively. 'Raw' denotes whether the CT images in the dataset are pre-processed. 'Image Size' refers to the CT image matrix size in the direction of the x, y, and z axes. 'Spacing' denotes the voxel sizes (mm) in the direction of the x, y, and z axes.

Dataset	Year	Scans	Class	Raw	Image Size	Spacing
LUNA16 [24]	2016	601	2	Yes	$512 \times 512 \times 95 - 512 \times 512 \times 733$	$(0.86, 0.86, 2.50) - (0.64, 0.64, 0.50)$
tianchi [29]	2017	800	2	Yes	$512 \times 512 \times 114 - 512 \times 512 \times 1034$	$(0.66, 0.66, 2.50) - (0.69, 0.69, 0.30)$
russia [19]	2018	364	2	Yes	$512 \times 512 \times 313 - 512 \times 512 \times 1636$	$(0.62, 0.62, 0.80) - (0.78, 0.78, 0.40)$
PN9 [18]	2021	8796	9	No	$212 \times 212 \times 181 - 455 \times 455 \times 744$	$(1.00, 1.00, 1.00) - (1.00, 1.00, 1.00)$

Table 2. Distribution of the pulmonary nodule size. 'd' indicates the nodule diameter (mm).

Dataset	$d < 3$	$3 \le d < 5$	$5 \le d < 10$	$10 \le d < 30$	$30 \le d$	All
LUNA16 [24]	-	270	635	279	2	1186
tianchi [29]	1	213	596	423	11	1244
russia [19]	6	552	907	360	25	1850
PN9 [18]	9	4678	29213	6053	483	40436

the CT images according to the extracted lung masks. Figure 3 shows the CT image samples after being pre-processed. For PN9 [18] dataset, the data pre-processing procedure is kept the same as in [18]. In our experiments, the voxel

coordinates are utilized. Based on our pre-processing procedures and the voxel coordinates, the nodule locations in the annotation files are recalculated.

In terms of the evaluation metric, the Free-Response Receiver Operating Characteristic (FROC), a commonly used measure for pulmonary nodule detection, is selected. It is calculated by averaging the sensitivities at 0.125, 0.25, 0.5, 1, 2, 4, and 8 false positives per scan. We also use the detection sensitivity at 8 false positives per image for evaluation, since false positives in the medical field are preferable to false negatives. The detected nodule is counted as a true positive if there exists one annotated nodule, and the distance between the center points of the detected nodule and the annotated nodule is smaller than the radius R of the annotated nodule. Otherwise, the detected nodule is considered a false positive.

3.2   Implementation Details

In our experiments, we employ the same backbone as the SANet [18], thus utilizing the weights pre-trained on PN9 [18] for source model training. Concretely, the backbone is U-shaped [22], consisting of a 3D ResNet50 [4] equipped with Slice Grouped Non-local modules [18] and a decoder. Different from [18], the backbone is followed by FPN [15] as neck, and the FCOS-style [28] anchor-free head for classification and localization. The network is optimized using the Stochastic Gradient Descent (SGD). The training batch size of the 3D patches is 16. We implement the patch-based input strategy for training and use the complete 3D volume for inference as in [18]. The learning rate, the momentum, and the weight decay coefficients are respectively fixed at 0.001, 0.9, and 1×10^{-4}. To obtain the source model, the network is set to be trained for a maximum of 30 epochs. For learning in the target domain, we tune the pre-trained source model for 1 epoch. For other training and testing hyper-parameters, we follow the [28], and specialize some hyper-parameters in the task of detecting pulmonary nodules. We use FPN [15] with two levels, a detection head with two classification/regression towers, and a radius of 3. All the experiments are carried out with PyTorch on four NVIDIA GeForce RTX 3090 GPUs, each having 24 GB of memory.

3.3   Results

We evaluate the proposed method by contrasting it with the baseline approach, which simply fine-tunes all the parameters of the source model using the labeled samples from the target domain. Experiments are conducted with 20%, 40%, 60%, 80%, and 100% labeled training samples from the target domains respectively, and the results are reported for the whole target testing sets. Table 3 lists the experimental results of our method and the baseline on target dataset LUNA16 [24] and tianchi [29]. Our method obviously outperforms the baseline. Meanwhile, it adapts more efficiently. It is especially noteworthy that utilizing only the feature extraction module adaptation, the first step of our method without the use of any labeled training samples from the target domain, already brings a good performance. This shows the potential of our method in the more
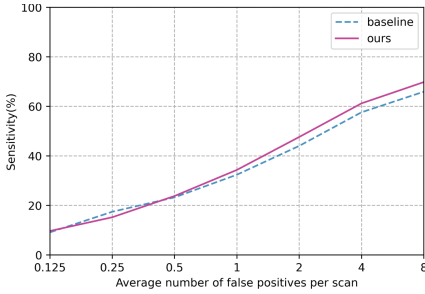
Fig. 4. FROC curves of our method and the baseline on target dataset russia w.r.t 60% percentage of its training set.

Fig. 5. FROC curves of our method and the baseline on target dataset russia w.r.t 80% percentage of its training set.

Table 3. Comparison of our method and the baseline on target dataset LUNA16 and tianchi w.r.t percentage of their training set. The values are pulmonary nodule detection sensitivities (unit: %) at 8 false positives per CT image, with each column indicating the percentage of the training set.

Method	LUNA16					tianchi				
	20%	40%	60%	80%	100%	20%	40%	60%	80%	100%
Fine-tuning	88.53	90.36	91.74	92.66	92.20	91.04	94.02	91.79	90.67	92.91
Ours (Step1)	85.77	89.90	88.99	89.44	88.99	93.28	93.28	93.28	92.91	93.28
Ours (+ Step2)	89.90	86.23	92.20	93.57	94.03	91.41	92.53	92.91	93.65	93.65

Table 4. Comparison of our method and the baseline on target dataset russia w.r.t percentage of its training set. The values are FROCs (unit: %) with each column indicating the percentage of the training set.

Method	russia				
	20%	40%	60%	80%	100%
Fine-tuning	**31.18**	**33.12**	35.69	36.52	38.26
Ours	29.91	33.08	**37.39**	**38.02**	**38.34**

wild and challenging settings. Nonetheless, the performance of our method on target dataset russia [19] is unsatisfactory, probably due to its larger shift with the source. For more adaptation, we tune all the parameters of the model in our second step on russia [19]. As listed in Table 4, our method obtains better FROC scores for lung nodule detection than the baseline, which verifies the effectiveness of our proposed adapting via entropy minimization. The FROC curves illustrated in Fig. 4 and Fig. 5 further confirm the superiority of our method.

4   Related Works

Recently, some works propose to adapt the trained model in test-time. This branch of study originates from the works of recalculating the batch statistics [14]. Test-time training (TTT) [26] relies on a proxy task for altering training the entire model on the source, and then adapts to the target using self-supervised learning. Tent [31] optimizes the affine parameters of batch normalization layers of the model via entropy minimization. This is demonstrated to be effective for robustness and source-free domain adaptation tasks. In [36], the authors replace the target statistics used in Tent with mixed source and target statistics. T3A [10] utilizes centroid-based modification to adapt the classifier in test-time for domain generalization. In [35], the authors revisit the batch normalization in the training process and develop a test-time batch normalization layer design named GpreBN, which is optimized during testing by minimizing entropy loss. This newly designed batch normalization operation preserves the same gradient backpropagation form as training and uses dataset-level statistics for robust optimization and inference. Unfortunately, all these works focus on image classification or semantic segmentation [11], and may not work well on object detection. In contrast, our method revisits the batch statistics for cross-domain pulmonary nodule detection, delving into the model optimization method specific for the detection.

5   Conclusion

In this paper, we present a source data-free setting for cross-domain lung nodule detection and present a method to tackle this issue, requiring only a pre-trained source model and a limited number of annotated samples from the target domain. Specifically, our method adapts the feature extraction module of the model by minimizing the proposed general entropy loss, and tunes the detection head with labeled target samples to enhance the detection performance even more. Experiments on our established benchmark verify that our method is an effective way to solve cross-domain object detection with data privacy issues involved. To the best of our knowledge, this is the first work on cross-domain pulmonary nodule detection without access to the source data. We also hope that this work in the medical field can bring insights into the general object detection field. In the future, we plan to pursue adaptation to more and harder types of shifts.

References

1. Cai, Q., Pan, Y., Ngo, C., Tian, X., Duan, L., Yao, T.: Exploring object relation in mean teacher for cross-domain detection. In: CVPR, pp. 11457–11466. Computer Vision Foundation/IEEE (2019)

2. Chen, Y., Li, W., Sakaridis, C., Dai, D., Gool, L.V.: Domain adaptive faster R-CNN for object detection in the wild. In: CVPR, pp. 3339–3348. Computer Vision Foundation/IEEE (2018)
3. Girshick, R.B.: Fast R-CNN, In: ICCV. pp. 1440–1448. IEEE (2015)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778. IEEE (2016)
5. He, M., et al.: Cross domain object detection by target-perceived dual branch distillation. In: CVPR, pp. 9560–9570. IEEE (2022)
6. He, Y., Zhu, C., Wang, J., Savvides, M., Zhang, X.: Bounding box regression with uncertainty for accurate object detection. In: CVPR, pp. 2888–2897. Computer Vision Foundation/IEEE (2019)
7. He, Z., Zhang, L.: Domain adaptive object detection via asymmetric tri-way faster-RCNN. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12369, pp. 309–324. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58586-0_19
8. Hofmanninger, J., Prayer, F., Pan, J., Rohrich, S., Prosch, H., Langs, G.: Automatic lung segmentation in routine imaging is a data diversity problem, not a methodology problem. CoRR abs/2001.11767 (2020)
9. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: ICML, vol. 37, pp. 448–456 (2015)
10. Iwasawa, Y., Matsuo, Y.: Test-time classifier adjustment module for model-agnostic domain generalization. In: NeurIPS, pp. 2427–2440 (2021)
11. Jiang, Y., et al.: A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation. IEEE ACM Trans. Comput. Biol. Bioinform. **18**(1), 40–52 (2021)
12. Khodabandeh, M., Vahdat, A., Ranjbar, M., Macready, W.G.: A robust learning approach to domain adaptive object detection. In: ICCV, pp. 480–490. IEEE (2019)
13. Li, X., et al.: Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection. In: NeurIPS (2020)
14. Li, Y., Wang, N., Shi, J., Liu, J., Hou, X.: Revisiting batch normalization for practical domain adaptation. In: ICLR (2017)
15. Lin, T., Dollár, P., Girshick, R.B., He, K., Hariharan, B., Belongie, S.J.: Feature pyramid networks for object detection. In: CVPR, pp. 936–944. IEEE (2017)
16. Lin, T., Goyal, P., Girshick, R.B., He, K., Dollár, P.: Focal loss for dense object detection. In: ICCV, pp. 2999–3007. IEEE (2017)
17. Liu, W., et al.: SSD: single shot multibox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2
18. Mei, J., Cheng, M.M., Xu, G., Wan, L.R., Zhang, H.: SANet: a slice-aware network for pulmonary nodule detection. IEEE Trans. Pattern Anal. Mach. Intell. **44**, 4374–4387 (2021)
19. Morosov, S., et al.: Tagged results of lung computed tomography scans (RU 2018620500) (2018)
20. Qiu, H., Li, H., Wu, Q., Shi, H.: Offset bin classification network for accurate object detection. In: CVPR, pp. 13185–13194. Computer Vision Foundation/IEEE (2020)
21. Redmon, J., Divvala, S.K., Girshick, R.B., Farhadi, A.: You only look once: unified, real-time object detection. In: CVPR, pp. 779–788. IEEE (2016)
22. 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

23. Saito, K., Ushiku, Y., Harada, T., Saenko, K.: Strong-weak distribution alignment for adaptive object detection. In: CVPR, pp. 6956–6965. Computer Vision Foundation/IEEE (2019)

24. Setio, A.A.A., et al.: Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. Med. Image Anal. **42**, 1–13 (2017)

25. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**(3), 379–423 (1948)

26. Sun, Y., Wang, X., Liu, Z., Miller, J., Efros, A.A., Hardt, M.: Test-time training with self-supervision for generalization under distribution shifts. In: ICML, vol. 119, pp. 9229–9248. PMLR (2020)

27. Tang, H., Zhang, C., Xie, X.: NoduleNet: decoupled false positive reduction for pulmonary nodule detection and segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11769, pp. 266–274. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_30

28. Tian, Z., Shen, C., Chen, H., He, T.: FCOS: fully convolutional one-stage object detection. In: ICCV, pp. 9626–9635. IEEE (2019)

29. Tianchi: Tianchi medical AI competition: Intelligent diagnosis of pulmonary nodules (2017). https://tianchi.aliyun.com/competition/entrance/231601/introduction

30. Tychsen-Smith, L., Petersson, L.: Improving object localization with fitness NMS and bounded IOU loss. In: CVPR, pp. 6877–6885. Computer Vision Foundation/IEEE (2018)

31. Wang, D., Shelhamer, E., Liu, S., Olshausen, B.A., Darrell, T.: TENT: fully test-time adaptation by entropy minimization. In: ICLR (2021)

32. Xu, C., Zhao, X., Jin, X., Wei, X.: Exploring categorical regularization for domain adaptive object detection. In: CVPR, pp. 11721–11730. Computer Vision Foundation/IEEE (2020)

33. Xu, R., et al.: SGDA: towards 3D universal pulmonary nodule detection via slice grouped domain attention. IEEE/ACM Trans. Comput. Biol. Bioinform. 1–13 (2023). https://doi.org/10.1109/TCBB.2023.3253713

34. Xu, R., Luo, Y., Du, B., Kuang, K., Yang, J.: LSSANet: a long short slice-aware network for pulmonary nodule detection. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13431, pp. 664–674. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16431-6_63

35. Yang, T., Zhou, S., Wang, Y., Lu, Y., Zheng, N.: Test-time batch normalization. CoRR abs/2205.10210 (2022)

36. You, F., Li, J., Zhao, Z.: Test-time batch statistics calibration for covariate shift. CoRR abs/2110.04065 (2021)

37. Zhang, S., Chi, C., Yao, Y., Lei, Z., Li, S.Z.: Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In: CVPR, pp. 9756–9765. IEEE (2020)

38. Zhang, Y., Wang, Z., Mao, Y.: RPN prototype alignment for domain adaptive object detector. In: CVPR, pp. 12425–12434. Computer Vision Foundation/IEEE (2021)

39. Zhao, G., Li, G., Xu, R., Lin, L.: Collaborative training between region proposal localization and classification for domain adaptive object detection. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12363, pp. 86–102. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58523-5_6