





Domain Generalization for Mammography Detection via Multi-style and Multi-view Contrastive Learning

Zheren Li^{1,2} , Zhiming Cui^{1,3}, Sheng Wang^{1,2}, Yuji Qi⁴, Xi Ouyang^{1,2},
Qitian Chen¹, Yuezhi Yang³, Zhong Xue¹, Dinggang Shen^{1,5},
and Jie-Zhi Cheng¹ 

¹ Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China
jzcheng@ntu.edu.tw

² Institute for Medical Imaging Technology, School of Biomedical Engineering,
Shanghai Jiao Tong University, Shanghai, China

³ Department of Computer Science, The University of Hong Kong,
Pok Fu Lam, Hong Kong

⁴ Department of Biomedical Engineering, Yale University, New Haven, USA

⁵ School of BME, ShanghaiTech University, Shanghai, China

Abstract. Lesion detection is a fundamental problem in the computer-aided diagnosis scheme for mammography. The advance of deep learning techniques have made a remarkable progress for this task, provided that the training data are large and sufficiently diverse in terms of image style and quality. In particular, the diversity of image style may be majorly attributed to the vendor factor. However, the collection of mammograms from vendors as many as possible is very expensive and sometimes impractical for laboratory-scale studies. Accordingly, to further augment the generalization capability of deep learning model to various vendors with limited resources, a new contrastive learning scheme is developed. Specifically, the backbone network is firstly trained with a multi-style and multi-view unsupervised self-learning scheme for the embedding of invariant features to various vendor-styles. Afterward, the backbone network is then recalibrated to the downstream task of lesion detection with the specific supervised learning. The proposed method is evaluated with mammograms from four vendors and one unseen public dataset. The experimental results suggest that our approach can effectively improve detection performance on both seen and unseen domains, and outperforms many state-of-the-art (SOTA) generalization methods.

Keywords: Domain generalization · Breast lesion detection · Contrastive learning

Z. Li and Z. Cui—Equal contribution.

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-87234-2_10) contains supplementary material, which is available to authorized users.

1 Introduction

The advance of deep learning (DL) techniques have remarkably improved the computer-aided detection (CAdE) of breast lesions in mammography. It has been shown in many studies [10, 12, 15] that the incorporation of DL-based CAdE software in the reading workflow of mammography can effectively improve the detection accuracy. Promising DL-based CAdE performance requires large and diverse training data. In particular, the inclusion of wide variety of vendors is very important to equip the DL-based CAdE system with prominent generalization capability. As shown in Fig. 1, the styles of images from various vendors vary significantly. Accordingly, the generalization of DL-based CAdE system may be limited if the data of each vendor are not sufficiently included in the training stage. However, the collection of large and diverse data with various vendors is very expensive and is impractical. Meanwhile, it is also well-known that there exists domain gap between datasets from various institutes. Therefore, a domain generalization method is needed to alleviate the demands of large and diverse data from various vendors for the DL-based CAdE scheme.

In the literature, the domain generalization for DL technique can be classified into three categories: 1) conventional data argumentation methods, e.g. rotation, flip, deformation and jittering [14, 18, 23], 2) learning based methods with the generative deep neural networks [6, 19, 21, 22], 3) learning based methods for the exploration of task-specific and domain-invariant features [2, 4, 5, 7–9, 20]. However, all these learning based methods relied on annotated data, which are costly obtain and availability is relatively limited. A new method to automatically mine domain-invariant features from large unannotated data is very needed.

In this study, to address the above mentioned issues, we explore the contrastive learning technique to augment the generalization capability of DL-based CAdE in self-supervised fashion. The contrastive learning has also been shown to generate better pretrained models for several medical image problems, e.g., diagnosis of chest radiography [1, 16, 25], dermatology images [1], etc., but less exploited for extracting domain invariant features. To specifically address the issue of vendor domain gap, we propose a multi-style and multi-view contrastive learning method to boost the lesion detection performance in mammography. Specifically, to attain the goal of generalization robustness to multiple vendor styles, the CycleGAN [26] technique is employed to generate multiple vendor-style images from a single vendor style image. The generated multiple style images from the same source are randomly paired as positive samples for the multi-style contrastive learning. Meanwhile, for the multi-view contrastive learning, the CC and MLO views of the same breast are also paired as positive samples. After self-supervised training, the backbone of the contrastive learning model is employed for the downstream detection task.

To our best knowledge, this is the first study that explores the contrastive learning to explicitly address the issue of vendor-style domain gap. We specifically train our method on three vendor domains, and then test on images from both seen and unseen domains. The results demonstrated that our method can effectively boost the lesion detection performance on either seen or unseen

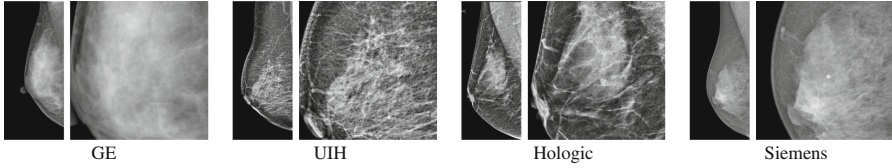


Fig. 1. Style differences among four mammography vendors, i.e., GE, United Imaging Healthcare(UIH), Hologic, and Siemens. All images of the four vendors are MLO views. To illustrate the detailed image variation, a zoom-in image is provided on the right of each MLO view.

domain. We also compare our multi-style and multi-view contrastive learning to several state-of-the-art domain generalization methods and suggest our pre-trained model can yield better generalization effect on both seen and unseen data.

2 Method

Figure 2 illustrates our domain generalization framework for lesion detection in mammography. We first conduct multi-style and multi-view contrastive learning scheme to extract domain invariant features on the data from different vendors. Afterward, the pretrained network backbone is transferred to the downstream task of lesion detection, especially on the data from unseen vendors.

2.1 Contrastive Learning Scheme

Contrastive learning is a self-supervised learning method, which trains an encoder to quantify the image representation into a proper vector space without the supervision of annotations. The derived model from contrastive learning can be served as a pretrained model for various downstream tasks like segmentation, detection, etc. The major advantage may lie in the flexibility to the downstream tasks.

The basic idea of contrastive learning is to bundle diversified images of the same class/object/subject as positive pairs for the exploration of reliable representation in feature space by the self-supervised learning manner. Specifically, given a mini-batch of N images, each example is randomly augmented twice with the diversifying operations, e.g., cropping, rotation, style-transferring, etc., to generate an augmented mini-batch with $2N$ samples. In the augmented mini-batch, two samples from the same image source are treated as a positive pair (i, j) , whereas the other $2(N - 1)$ samples within the mini-batch are regarded as negative pairs. With the positive and negative pairs, the contrastive learning is driven with the contrastive loss to maximize the agreement for the positive pairs. The contrastive loss is defined as:

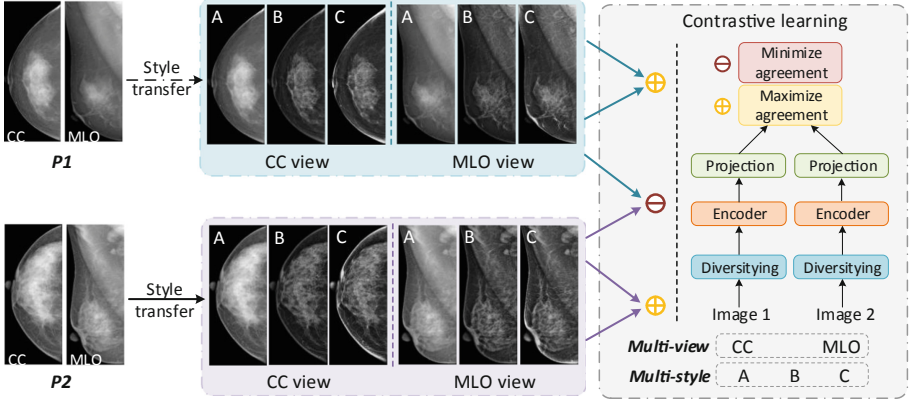


Fig. 2. Illustration of multi-style and multi-view contrastive learning.

$$\ell_{i,j} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\text{sim}(z_i, z_k)/\tau)}, \quad (1)$$

where $\text{sim}(\cdot)$ is the dot product and z refers to the extracted features. $\mathbb{1}_{[k \neq i]} \in \{0, 1\}$ is an indicator function equaling 1 when $k \neq i$ and τ is a temperature parameter.

Via maximizing the agreement for the positive pairs, the learnt features are supposed to be more general and robust to the diversifying operations. We further leverage the concept of contrastive learning to explore the generalization for various vendor styles and the domains of CC and MLO views. The details of multi-style and multi-view contrastive learning are elaborated in the following section.

2.2 Multi-style and Multi-view Contrastive Learning

Multi-style Contrastive Learning. Referring to Fig. 2, the image styles vary to different vendors and are regarded as distinctive domains. To endow the backbones of the lesion detectors better vendor-style generalization capability, we exploit the CycleGAN [26] technique as the diversifying operation for the contrastive learning. Specifically, given M seen vendor-style domain, we train $\binom{M}{2}$ generators, which map the data distribution of source domain Ω_i to target Ω_j domain, $\forall i, j \in M$. Comparing to the method used in [19], the CycleGAN realizes the style transfer with bidirectional learning process. The work [19] unidirectionally takes a few references images may attain limited transferring effect.

With the $\binom{M}{2}$ generators, each of the original N images in the training set can be diversified into $M - 1$ transferred images. Afterward, a mini-batch with $M \times N$ images can be formed. The positive pairs for the contrastive learning are constituted with any two images of the same image source in the original N image set. Therefore, there are possible $N \times \binom{M}{2}$ positive pairs that be randomly

selected in the contrastive learning, whereas the other combination pairs are treated as negative pairs. The multi-style contrastive learning is then carried with minimizing the Eq. 1 to seek feature embedding space with better generalization capability to various vendor domains.

Multi-view Contrastive Learning. A standard examination of mammography consists of two CC and MLO views. Because the two standard views are mutually complementary, the appearance of CC and MLO images are different. For example, a MLO view includes axilla region, while a CC view doesn't. Accordingly, domain gap between these two views may exist. To seek view domain-invariant feature embedding for better lesion detection performance, we explore contrastive learning scheme to consider the distinctive view domains. Specifically, we treat the CC and MLO view of the same breast from the same patient as positive pair, whereas the other combination of CC and MLO is a negative pair. To further enrich the sample diversity, we further implement diversifying operations including random cropping, random rotation, horizontal flipping, and adjustment of brightness, contrast, and saturation for the CC and MLO in each positive or negative pair. With the prepared sample pairs, the multi-style contrastive learning can be carried out for the embedding of view-invariant features.

The synergy of the multi-style and multi-view contrastive learning is realized as illustrated in Fig. 2. Specifically, we firstly employ the style transferring CycleGAN to derive diversified vendor styles and perform the random diversifying operations on the CC and MLO views as the sample pairs for contrastive learning. We will show that our multi-style and multi-view contrastive learning can boost the lesion detection tasks effectively on either seen and unseen domains.

2.3 Lesion Detection Detection with Contrastive Learning

In this study, we employ the classic detection network of FCOS [17] to identify mass and clustered calcifications in mammography. The derived pretrained model from the contrastive learning is adopted as the backbone of the FCOS architecture for the realization of the downstream lesion detection tasks.

3 Experiments

Implementation Details. Our method was developed based on Pytorch package and trained with NVIDIA Titan RTX GPUs. We train CycleGAN [24, 26] models with the settings of 50, 100, 200 epochs, and choose the setting, i.e., 100 epochs, that can achieve best lesion detection performance on the validation set. The backbone of the generator is ResNet with 9 blocks with 20 conv layers, while the discriminator is PatchGAN with 6 conv layers for binary classifications. In training, the sizes of inputs and outputs are $512 * 512$ by random cropping the images. MSE and L1 losses are used for classification and reconstruction. For a

style transfer, e.g., A to B, the involved training data from A and B are both 1000 for balanced training. Code and models of style transfer are available at: https://github.com/lizheren/MSVCL_MICCAI2021.

We adopt ResNet-50 as the backbone model for contrastive learning and FCOS detector. For fair comparison, the learning rate and batch size for all contrastive learning schemes are set the same as 0.3 and 256, respectively. Meanwhile, all contrastive learning schemes in all experiments use the same diversifying operations, including random cropping, random rotation in $\pm 10^\circ$, horizontal flipping and random color jittering (strength = 0.2). For the training of FCOS models, the SGD method is adopted with the parameters of learning rate, weight decay and momentum set as 0.005, 10^{-4} , 0.9, respectively. The epoch and batch size are set to 50 and 8, respectively, throughout all experiments. Several augmentation methods, e.g. random flipping, scaling, etc., are also implemented in the training of FCOS.

3.1 Datasets

In this paper, mammograms of four vendors, i.e., GE, United Imaging Healthcare(UIH), Hologic, and Siemens, denoted as A, B, C, and D, respectively, are involved. All the data of the four vendors were collected from Asian women. The data of vendors A, B and C are set as seen domains whereas the vendor D is treated as unseen domain. To evaluate the generalization capability of our method, we also involve the public dataset, INbreast [13], denoted as E as unseen dataset. In total, this study involves 28,700 mammograms, where 27,000 unannotated images are used for style transfer and contrastive learning. The remaining 1,700 annotated images are adopted for the training/validation/testing of the detection tasks. We conduct a preprocessing step to align various mammograms from A, B, C, D, and E into the same pixel spacing of 0.1 mm to facilitate the training of deep learning methods. For the assessment of detector, we adopt the mean average precision(mAP) metrics for the quantitative comparison (Table 1).

Table 1. Details of usage for the data involved in this study. Specifically, the columns “Style Transfer” and “Self-Supervision” suggest the number of images involved in the training of CycleGAN and contrastive learning, respectively.

Domain	Dataset	Style transfer	Self-supervision	Detection (train/val/test)	Vendor
Seen	Style A	1000	8000	360/40/100	GE
	Style B	1000	8000	360/40/100	UIH
	Style C	1000	8000	360/40/100	HOLOGIC
Unseen	Style D	0	0	0/0/100	SIEMENS
	Style E	0	0	0/0/100	INbreast

3.2 Ablation Study

The ablation experiments are composed of two parts. First, we evaluate efficacy of original simple contrastive learning (SimCLR) [3] on the lesion detec-

tion tasks. Specifically, we compare lesion detection performance with 1) no pretraining, 2) ImageNet, 3) SimCLR on MammoPre, and 4) SimCLR on ImageNet \rightarrow MammoPre pretrained models and report the corresponding performance in the first to fourth rows of Table 2. The third row suggests that SimCLR is trained from scratch with the unlabeled images of vendors A, B and C, where these unlabeled set is denoted as MammoPre. The fourth row ImageNet \rightarrow MammoPre indicates that SimCLR is initialized with ImageNet parameters and then trained with the MammoPre set. In this part, the pretrained model of ImageNet \rightarrow MammoPre can yield better detection performance.

The second part is to assess the effectiveness of each component of our framework. Specifically, we compare the pretrained models derived from multi-style contrastive learning (MSCL), multi-view contrastive learning (MVCL) and the combination of multi-style and multi-view contrastive learning (MSVCL) w.r.t. the lesion detection performances, which are listed in the fifth to seventh rows in Table 2, respectively. As can be found in Table 2, the effectiveness of our contrastive learning for the lesion detection tasks on both seen and unseen domains is corroborated. The detailed ablation analysis w.r.t. the mass and clustered calcifications detection tasks can be found in the Tables 1 and 2 of the supplement.

Table 2. Ablation analysis w.r.t. the pre-training strategy and our methods. The results on mAP of the mass and clustered calcifications detection tasks are reported. The performances w.r.t. the mass and clustered calcifications detection tasks are shown in the Table 1 and Table 2 of the Supplement, respectively.

Method	Pretrain strategy	Seen domain				Unseen domain		
		Style A	Style B	Style C	Avg.	Style D	Style E	Avg.
Random	None	0.427	0.563	0.487	0.492	0.345	0.525	0.435
Supervised	ImageNet	0.731	0.7805	0.728	0.746	0.673	0.811	0.742
SimCLR	MammoPre	0.740	0.772	0.742	0.751	0.665	0.794	0.729
SimCLR	ImageNet \rightarrow MammoPre	0.749	0.774	0.759	0.761	0.687	0.819	0.753
MSCL	ImageNet \rightarrow MammoPre	0.768	0.780	0.775	0.774	0.703	0.845	0.774
MVCL	ImageNet \rightarrow MammoPre	0.756	0.785	0.771	0.771	0.706	0.830	0.768
MSVCL	ImageNet \rightarrow MammoPre	0.779	0.812	0.784	0.792	0.717	0.862	0.789

3.3 Comparison with State-of-the-Art (SOTA) Methods

To further compare with other domain generalization methods, three SOTA methods, i.e., BigAug [23], Domain Diversification(DD) [6], and EISNet [20] are implemented. The BigAug [23] is a conventional data augmentation method, whereas the DD [6] proposed a generative learning method for domain generalization. The EISNet [20] is a learning-based method that explores task-specific and domain-invariant features. Our method on the other hand can decouple the downstream tasks intrinsically and provide task- and domain-invariant features. We trained all methods with the multi-view images from various vendors and explored the domain generalization correspondingly.

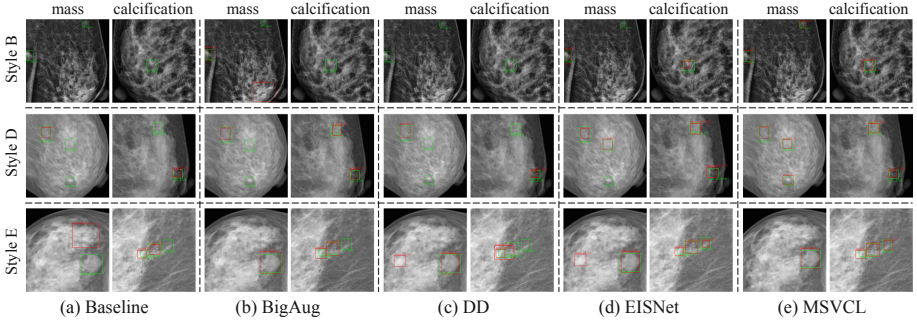


Fig. 3. Qualitative results for mass and clustered calcifications detection from different methods on one seen domain (style B) and two unseen domains (style D, E). The green boxes refer to the lesion regions labeled by the radiologist. The red boxes refer to the lesion regions detected by different methods. (Color figure online)

The comparison results are shown in Table 3, where the Baseline row suggests the result of SimCLR with ImageNet \rightarrow MammoPre training strategy. As it can be found, our MSVCL can promise the best lesion detection performance on either seen or unseen domains. In particular, our method outperformed the EISNet by 1.8% on seen domains and 2.2% on unseen domains and data. Therefore, our MSVCL can be a new referential method for the pretraining of useful backbone for the downstream mammographic image analysis problems. The details performance comparison w.r.t. the mass and clustered calcifications detection tasks can be found in the Tables 3 and 4 of the supplement.

Table 3. Performance comparison between our MSVCL and other SOTA domain generalization methods. The results on mAP of the mass and clustered calcifications detection tasks are reported. The performances w.r.t. the mass and clustered calcifications detection tasks are shown in the Table 3 and Table 4 of the Supplement, respectively.

Method	Seen domain				Unseen domain		
	Style A	Style B	Style C	Avg.	Style D	Style E	Avg.
Baseline	0.749	0.774	0.759	0.761	0.687	0.819	0.753
BigAug [23]	0.756	0.792	0.743	0.763	0.692	0.829	0.760
DD [6]	0.754	0.792	0.770	0.772	0.695	0.821	0.758
EISNet [20]	0.765	0.781	0.775	0.774	0.690	0.845	0.767
MSVCL (ours)	0.779	0.812	0.784	0.792	0.717	0.862	0.789

To further visually illustrate the efficacy of our MSVCL pretrained model, the t-SNE [11] is employed to visualized data distribution of various vendor domains in the embedded feature space. Figure 4 compares the pretrained models of Baseline and MSVCL. As can be found, our MSVCL can better break the

distribution boundaries between various vendor domains. Therefore, the encoded features shall be more style-invariant. We also demonstrate the detection results visually in the Fig. 3 and the style transfer results of CycleGAN in Fig. 1 of the supplement for more visual assessment.

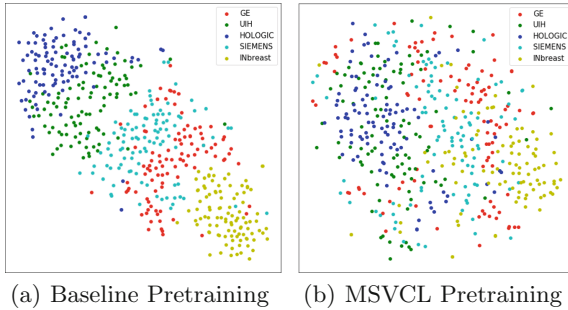


Fig. 4. t-SNE visualization for the pre-trained models from Baseline and MSVCL.

4 Conclusion

A new domain generalization method is proposed to assist the lesion detection schemes in mammography. Specifically, we conduct multi-style and multi-view contrastive learning scheme to embed domain-invariant features to various vendors. The experimental results suggest that our domain generalization method can help to significantly improve the lesion detection tasks on both seen and unseen domains. In particular, for unseen domain of INbreast dataset, we also can achieve the best performance, compared to the three implemented SOTA domain generalization methods. Therefore, the efficacy of our method is corroborated.

References

1. Azizi, S., et al.: Big self-supervised models advance medical image classification. arXiv preprint [arXiv:2101.05224](https://arxiv.org/abs/2101.05224) (2021)
2. Chen, N., et al.: Unsupervised learning of intrinsic structural representation points. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9121–9130 (2020)
3. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning, pp. 1597–1607. PMLR (2020)
4. Cui, Z., et al.: Structure-driven unsupervised domain adaptation for cross-modality cardiac segmentation. IEEE Trans. Med. Imaging (2021)
5. Dou, Q., Castro, D.C., Kamnitsas, K., Glocker, B.: Domain generalization via model-agnostic learning of semantic features. arXiv preprint [arXiv:1910.13580](https://arxiv.org/abs/1910.13580) (2019)

6. Kim, T., Jeong, M., Kim, S., Choi, S., Kim, C.: Diversify and match: a domain adaptive representation learning paradigm for object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12456–12465 (2019)
7. Li, H., Pan, S.J., Wang, S., Kot, A.C.: Domain generalization with adversarial feature learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5400–5409 (2018)
8. Li, Y., et al.: Deep domain generalization via conditional invariant adversarial networks. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 624–639 (2018)
9. Liu, Q., Dou, Q., Heng, P.-A.: Shape-aware meta-learning for generalizing prostate MRI segmentation to unseen domains. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12262, pp. 475–485. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59713-9_46
10. Lotter, W., et al.: Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. *Nat. Med.* **27**(2), 244–249 (2021)
11. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**(11) (2008)
12. McKinney, S.M., et al.: International evaluation of an AI system for breast cancer screening. *Nature* **577**(7788), 89–94 (2020)
13. Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S.: INbreast: toward a full-field digital mammographic database. *Acad. Radiol.* **19**(2), 236–248 (2012)
14. Romera, E., Bergasa, L.M., Alvarez, J.M., Trivedi, M.: Train here, deploy there: robust segmentation in unseen domains. In: 2018 IEEE Intelligent Vehicles Symposium (IV), pp. 1828–1833. IEEE (2018)
15. Salim, M., et al.: External evaluation of 3 commercial artificial intelligence algorithms for independent assessment of screening mammograms. *JAMA Oncol.* **6**(10), 1581–1588 (2020)
16. Sowrirajan, H., Yang, J., Ng, A.Y., Rajpurkar, P.: MoCo pretraining improves representation and transferability of chest x-ray models. *arXiv preprint arXiv:2010.05352* (2020)
17. Tian, Z., Shen, C., Chen, H., He, T.: FCOS: fully convolutional one-stage object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9627–9636 (2019)
18. Volpi, R., Namkoong, H., Sener, O., Duchi, J.C., Murino, V., Savarese, S.: Generalizing to unseen domains via adversarial data augmentation. In: NeurIPS (2018)
19. Wang, S., et al.: mr²NST: multi-resolution and multi-reference neural style transfer for mammography. In: Rekik, I., Adeli, E., Park, S.H., Valdés Hernández, M.C. (eds.) PRIME 2020. LNCS, vol. 12329, pp. 169–177. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59354-4_16
20. Wang, S., Yu, L., Li, C., Fu, C.-W., Heng, P.-A.: Learning from extrinsic and intrinsic supervisions for domain generalization. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12354, pp. 159–176. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58545-7_10
21. Yue, X., Zhang, Y., Zhao, S., Sangiovanni-Vincentelli, A., Keutzer, K., Gong, B.: Domain randomization and pyramid consistency: simulation-to-real generalization without accessing target domain data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2100–2110 (2019)

22. Zakharov, S., Kehl, W., Ilic, S.: DeceptionNet: network-driven domain randomization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 532–541 (2019)
23. Zhang, L., et al.: Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. *IEEE Trans. Med. Imaging* **39**(7), 2531–2540 (2020)
24. Zhang, Y., Miao, S., Mansi, T., Liao, R.: Task driven generative modeling for unsupervised domain adaptation: application to X-ray image segmentation. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11071, pp. 599–607. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_67
25. Zhou, H.-Y., Yu, S., Bian, C., Hu, Y., Ma, K., Zheng, Y.: Comparing to learn: surpassing ImageNet pretraining on radiographs by comparing image representations. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12261, pp. 398–407. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59710-8_39
26. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232 (2017)