Computers in Biology and Medicine 185 (2025) 109455

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

# Domain generalization for mammographic image analysis with contrastive learning

Zheren Li [a,b], Zhiming Cui [c], Lichi Zhang [b], Sheng Wang [b,c], Chenjin Lei [a], Xi Ouyang [a], Dongdong Chen [b], Xiangyu Zhao [b], Chunling Liu [d,e], Zaiyi Liu [d,e], Yajia Gu [f], Dinggang Shen [a,c], Jie-Zhi Cheng [a,*]

[a] Shanghai United Imaging Intelligence Co., Ltd., Shanghai 200030, China
[b] The School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[c] The School of Biomedical Engineering, ShanghaiTech University, Shanghai 201210, China
[d] The Department of Radiology, Guangdong Provincial People's Hospital, Southern Medical University, Guangzhou 510080, China
[e] Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application, Guangzhou 510080, China
[f] The Department of Radiology, Fudan University Shanghai Cancer Center, 270 Dongan Road, Shanghai 200032, China

ARTICLE INFO

*Keywords:*
Domain generalization
Mammographic image analysis
Contrastive learning

ABSTRACT

The deep learning technique has been shown to be effectively addressed several image analysis tasks in the computer-aided diagnosis scheme for mammography. The training of an efficacious deep learning model requires large data with diverse styles and qualities. The diversity of data often comes from the use of various scanners of vendors. But, in practice, it is impractical to collect a sufficient amount of diverse data for training. To this end, a novel contrastive learning, MSVCL+, is developed to equip the deep learning models with better style generalizability. Specifically, the multi-style and multi-view unsupervised self-learning scheme is carried out to seek robust feature embedding against style diversity as a pretrained model. Afterward, the pretrained network is further fine-tuned to the downstream tasks, e.g., mass detection, matching, BI-RADS rating, and breast density classification. The proposed method has been evaluated extensively and rigorously with mammograms from various vendor style domains and several public datasets. The experimental results suggest that the proposed domain generalization method can effectively improve performance of four mammographic image tasks on the data from both seen and unseen domains, and outperform many state-of-the-art (SOTA) generalization methods.

## 1. Introduction

Breast cancer is one of the leading causes of cancer-related deaths among women [1]. Mammography is the conventional imaging technique for early screening of breast cancer. It has been shown in many studies [2–5] that advances in deep learning (DL) techniques have remarkably improved the computer-aided detection and diagnosis (CAD) schemes of mammography. The CAD schemes are composed of several components, such as lesion detection [6], lesion matching [7], malignancy classification [8], density classification [9], and others [10]. However, most CAD schemes are not well tested for out-of-distribution generalization, i.e., the applicability to the unseen domains in the training process. Domain shift may result in a significant decline in performance across various vendors [11,12].

As shown in Fig. 1, the styles of images from various vendors vary significantly. The diversity in image style among mammograms from different vendors can be attributed to various factors, such as imaging hardware and processing pipeline. With different hardware settings, the following reconstruction and post-processing algorithms may need to be specially tailored and tuned. Therefore, the style of finally generated image can appear very distinctive from vendor to vendor.

These differences in image style pose a significant challenge for CAD systems, particularly those equipped with deep learning. This is because these systems require training data that are diverse and representative of the variability in the real world. However, it is extremely expensive and impossible to gather vast amounts of diverse data from numerous vendors. Meanwhile, it is well known that domain gaps exist between datasets from different hospitals, primarily due to the differences in institutional imaging conventions and protocols. Accordingly, the generalization of DL-based CAD systems may be limited if the data of each vendor is not sufficiently included in the training stage. To overcome

* Corresponding author.
*E-mail address:* jzcheng@ntu.edu.tw (J.-Z. Cheng).

https://doi.org/10.1016/j.compbiomed.2024.109455
Received 25 April 2024; Received in revised form 10 October 2024; Accepted 18 November 2024
Available online 9 December 2024

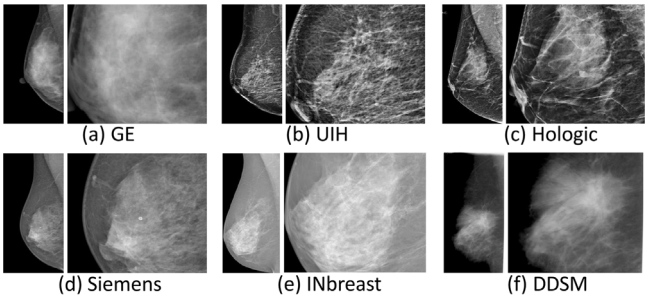


**Fig. 1.** Appearance differences across six styles/domains of mammograms: (a) GE, (b) United Imaging Healthcare (UIH), (c) Hologic, (d) Siemens, (e) INbreast, and (f) DDSM. For each style/domain, a zoomed-in of MLO view is provided to facilitate the visual comparison.

this challenge, a domain generalization method is needed to alleviate the burden of collecting large and diverse data from various vendors for DL-based CAD schemes.

To address the specific challenges in mammography analysis we propose a novel application of contrastive learning techniques to enhance the generalizability. While contrastive learning has shown promise in various medical imaging domains, such as chest radiography [13,14] and dermatology [13], its potential for extracting domain-invariant features in mammography remains largely unexplored. Our work specifically targets the vendor domain gap and view position variations unique to mammography. We introduce a multi-style and multi-view contrastive learning (MSVCL+) method designed to improve CAD performance across different vendor styles and view positions in mammography. To tackle the challenge of generalization across multiple vendor styles, we first employ CycleGAN [15] to generate diverse vendor-style images from a single source. These multi-style images are then used as positive samples in our multi-style contrastive learning (MSCL+) approach. Additionally, to address the variability between Craniocaudal (CC) and Mediolateral Oblique (MLO) views, we pair these views of the same breast as positive samples in our multi-view contrastive learning (MVCL+) scheme. This combined approach aims to create a more robust feature representation that is invariant to both vendor-specific image characteristics and view position differences. Following self-supervised training, we utilize the backbone of our contrastive learning model for various downstream tasks critical in mammography analysis. These tasks include mass detection, multi-view mass matching, BI-RADS rating, and breast density classification. For the multi-view mass matching task, we further employ contrastive learning by treating input Region of Interest (ROI) patches from the CC and MLO views of the same mass as positive pairs. This comprehensive approach addresses multiple aspects of the domain generalization problem specific to mammography, potentially improving the overall performance and reliability of CAD systems across different clinical settings and imaging equipments.

The domain generalization for the DL technique can be classified into three categories: (1) Conventional data augmentation methods, e.g., rotation, flip, deformation, and color jittering [16,17]. These augmentation methods can help the model adopt images from slightly different domains. (2) Learning-based methods with generative deep neural networks that synthesize data in the target domain [18–20]. These networks are implemented by learning an image-to-image mapping from the original to the target domain. (3) Learning-based methods for the exploration of task-specific and domain-invariant features [21–24]. Instead of learning an image-to-image mapping, these methods directly learn a representation-to-representation mapping. However, all of these learning-based approaches rely on labeled data which can be expensive to acquire and is often limited in availability. It is crucial to develop a novel approach for automatically extracting domain-invariant features from large, unlabeled datasets.

To address the above-mentioned issues, here we explore the contrastive learning technique to augment the generalizability of DL-based CAD. Contrastive learning has also been shown to generate better pre-trained models for several medical image problems, e.g., diagnosis of chest radiography [25,26] and dermatology images [25], but is less exploited for extracting domain-invariant features. To specifically address the issue of the vendor domain gap, MSVCL+ is proposed to boost the CAD performance in mammography. Specifically, to attain the goal of generalization robustness to multiple vendor styles, the CycleGAN [27] technique is employed to generate multiple vendor-style images from a single vendor-style image. The generated multi-style images from the same source are randomly paired as positive samples for MSCL+. To further gain generalizability for the position view, the CC and MLO views of the same breast are also paired as positive samples in the scheme of MVCL+. After self-supervised training, the backbone of the contrastive learning model is employed for the downstream tasks, including mass detection, multi-view mass matching, BI-RADS rating, and breast density classification. For the multi-view mass matching task, we also employ the contrastive learning approach in which the input region of interest (ROI) patches from the CC and MLO views of the same mass are treated as a positive pair. The "+" signs of MSVCL+, MSCL+, and MVCL+ are to distinguish them from the same notations in the previous work [15].

Our main contributions are summarized in threefold:

- A novel contrastive learning scheme is proposed to boost the generalizability and augment the robustness of the mammographic image analysis tasks. With style samples augmented by CycleGAN and style blending methods, the contrastive learning scheme can seek a robust feature embedding against various vendor styles. To our best knowledge, this is the first thorough study to explore the self-learning technique in addressing the mammographic domain gap problem w.r.t. vendor styles and views.
- The proposed method has been shown to be helpful in boosting the domain generalizability for the four mammographic tasks, namely mass detection, matching, BI-RADS rating, and breast density classification. The performances of the four tasks have been extensively and thoroughly evaluated on both seen and unseen domains.
- A relatively large dataset, 27,000 unlabeled and 2700 labeled images, was involved in the development and testing of the contrastive learning and the deep learning schemes for the mammographic tasks. The proposed method is trained on three seen vendor domains and evaluated on data from six vendor domains, including seen and unseen domains. In particular, two public datasets of unseen domains were collected from populations with ethics distinctive from the in-house data. Meanwhile, a rigorous data-hungry experiment is conducted to illustrate that the proposed domain generalization method may promise better task performance than other compared SOTA methods.

This study is an extension work of [15] with significant improvements. These improvements can be summarized in threefold. (1) Previous work [15] employed CycleGAN to synthesize discrete vendor styles. Therefore, the scope of vendor styles may be finite for contrastive learning. To further enrich the diversity of the vendor styles, a simple blending method is adopted to augment the variation of synthesized styles. Higher variation of styles may offer more search space for contrastive learning to seek better feature embedding. (2) In this study, the bundling strategy for the positive and negative pairs in contrastive learning is revised based on the knowledge of mammography. The strategy may better help the contrastive learning to find a more robust feature embedding. (3) The size of labeled data is expanded by 75%, encompassing a larger training and validation dataset, along with the addition of a new unseen testing dataset, while the proposed method is also subjected to more thorough and rigorous evaluation across three additional downstream tasks.

The rest of the paper is organized as follows: Section 2 gives a brief review of related works. Section 3 demonstrates the details of the proposed domain generalization method. Section 4 describes experimental results and visualizations. Finally, Section 5 concludes the paper.

## 2. Related works

### 2.1. Domain generalization

Domain generalization is an important research area to overcome performance degradation caused by cross-domain variations in medical image analysis. Many methods [14,22,23,28,29] have been proposed to improve generalizability through data augmentation or exploiting the general learning strategy. However, learning-based methods are more desirable and related to our work, we focus on this category in the following.

With the recent developments in machine learning techniques, plenty of works utilize representation learning to address the domain generalization problem. They can be mainly divided into four categories: kernel-based methods [30,31]; domain adversarial learning [32,33]; explicit feature alignment [34–36]; and invariant risk minimization [13,37]. Explicit feature alignment aims to align the features from source domains to learn domain-invariant representations. For example, Motiian et al. [34] develop a cross-domain contrastive loss for representation learning in which mapped domains are semantically matched, while remaining maximally separated. Pan et al. [35] implement instance normalization layers to CNNs to improve model generalization. Recently, Fan et al. [36] present that adaptively learning the normalization with the combination of multiple normalization strategies can improve domain generalization capabilities. The method [38] combines both the data argumentation and domain invariant feature extraction strategies to improve the domain generalization ability. The algorithms [39,40] first split the seen domain to various new domains, following by a model, e.g., meta-learning, to extract representative features.

Following the success of self-supervised learning [41], several studies [42–44] build self-supervised tasks from large-scale unlabeled data to learn generalized representations. For instance, Carlucci et al. [42] present a self-supervision task of solving jigsaw puzzles to learn the concepts of spatial correlation. Daehee et al. [43] propose a self-supervised contrastive regularization approach that utilizes only positive data pairs. Seogkyu et al. [44] exploit stylized features for regularization via consistency loss and domain-aware supervised contrastive loss. However, these methods were not specifically designed for medical images. Therefore, the efficacy of these methods [42–44] for medical image analysis is unknown.

### 2.2. Contrastive learning

Recently, contrastive learning has been proven to be notably effective in self-supervised learning. It draws samples from the same class (a positive pair) close together while drives different samples (or negative pairs) apart in the latent embedding space through contrastive loss. Chen et al. [45] propose SimCLR, a contrastive learning-based system for learning effective presentations by maximizing the similarity between two different augmented views from the original image. He et al. [46] utilize momentum contrast (MoCo) which divides each image into a query, and then generates a key by performing two distinct augmentations. MoCo v2 [47] incorporates enhancements such as the integration of a two-layer MLP head with ReLU during the unsupervised training stage, and the implementation of a data augmentation technique involving blurring. SimCLR and MoCo provide promising results that are much closer to supervised-learning compared to other self-supervised learning methods. These frameworks are widely used in the pre-train stage and showed constant improvement for various downstream tasks. Later, Grill et al. [48] introduce a technique called BYOL for learning feature representations without the massive number of negative pairs. Basically, BYOL added another MLP on the SimCLR to create an asymmetrical architecture. In this paper, we adopt the SimCLR as the pre-training method, because of its advantage of easy implementation. Meanwhile, we further replace the SimCLR with both MoCo v2 and BYOL as the pretraining methods. The experimental results suggest there is no significant difference among all pre-training methods.

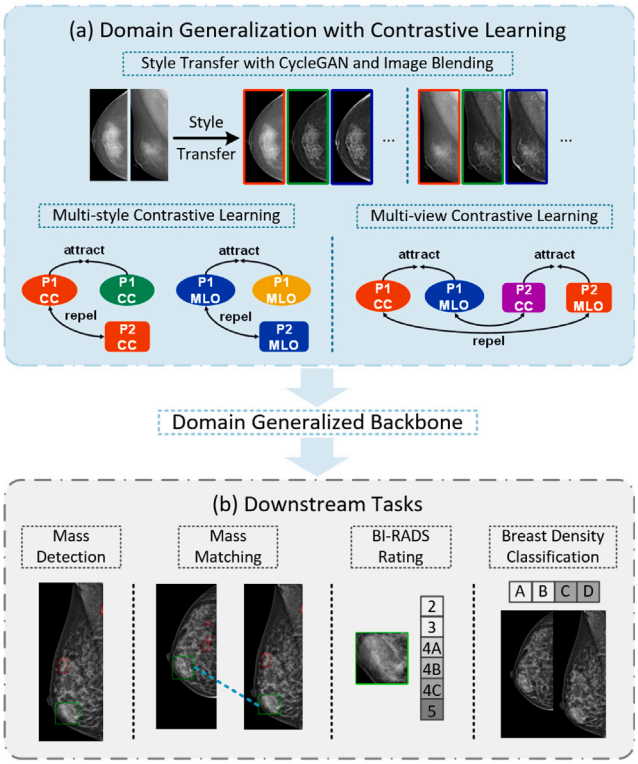


**Fig. 2.** Our domain generalization method for mammograms involves two main steps. In stage (a), the CycleGAN augmented with an image blending technique is adopted to generate diverse vendor styles. The contrastive learning scheme is further applied to learn better feature embedding against various domains, from the generated styles. In stage (b), the feature embedding encoded in the backbone of the contrastive learning model serves as the pre-trained model for downstream tasks.

### 2.3. Mammographic image analysis

Mass detection is one of the most fundamental problems in mammographic image analysis [6]. Jung et al. [49] use the RetinaNet [50] model as a one-stage mass detector in mammograms. Shen et al. [51] propose a framework that depends on adversarial learning to detect the mass in mammograms. The adversarial learning helps align the latent target features from unlabeled datasets with labeled source domain latent features. Moreover, CAD schemes also require supplementary mammographic image analysis functions such as mass matching, malignancy classification, and breast density classification. Yan et al. [7] exploit multi-tasking properties of deep networks to jointly learn mass matching and classification. Yang et al. [52] present MommiNet-v2 to incorporate the malignant information from both biopsies and BI-RADS categories. Zhao et al. [9] present an innovative bilateral-view adaptive spatial and channel attention network (BASCNet) for fully automated breast density classification. Despite the many existing methods for mammographic image analysis, few of them have been evaluated on unseen domains, which means that their out-of-distribution generalization has not been thoroughly examined.

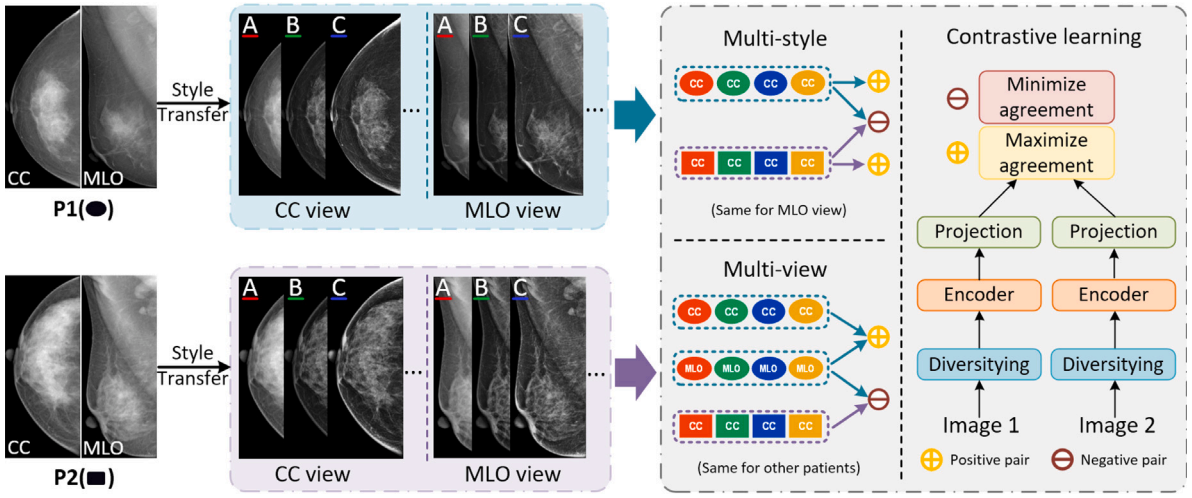


**Fig. 3.** Illustration of the MSVCL+ scheme. P1 and P2 indicate different patients. The letters A, B, and C represent different vendor styles generated from one source domain. The right box shows the concept of bundling positive and negative pairs for the contrastive learning schemes. Specifically, the red circle with a minus sign inside stands for the negative pair, while the yellow circle with a plus sign inside indicates the positive pair.

## 3. Method

Fig. 2 illustrates the proposed domain generalization framework for mammography analysis. Our approach consists of two main stages. In stage (a), a style-robust backbone is trained with the contrastive learning technique as the pre-trained model for the downstream tasks. To facilitate contrastive learning, the CycleGAN technique is first employed to diversify the vendor styles. In the following, a multi-style and multi-view contrastive learning scheme is carried out to embed a general feature space, which is more robust to both the vendor-style and view domains, in the pre-trained backbone. The common downstream tasks of mammography analysis, including mass detection, matching, BI-RADS rating, and breast density classification, are further fine-tuned with the pre-trained backbone for better generalizability.

### 3.1. Contrastive learning scheme

Contrastive learning is a self-supervised learning method that trains a network to encode the image representation into a proper vector space without the requirement of explicit annotations. The derived model from contrastive learning commonly serves as a pre-trained model for the training of various downstream tasks.

The basic idea of contrastive learning is to pack the diversified images of the same class/object/subject as positive pairs for the exploration of proper feature embedding. Specifically, given a mini-batch of $N$ images, each example is randomly augmented twice with diversifying operations, e.g., cropping and rotation, 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, contrastive learning is driven by the contrastive loss to maximize the agreement for the positive pairs. The contrastive loss is defined as:

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

where $\mathrm{sim}(\cdot)$ 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 $k \neq i$. $\tau$ is a temperature parameter.

By maximizing the agreement for the positive pairs, the learned features of corresponding images are supposed to "attract" each other, while the learned features of non-corresponding images "repel" each other by minimizing the agreement for the negative pairs. We further leverage the concept of contrastive learning to explore the generalization of various vendor styles and the domains of CC and MLO views. The details of MSVCL+ are elaborated in the following section.

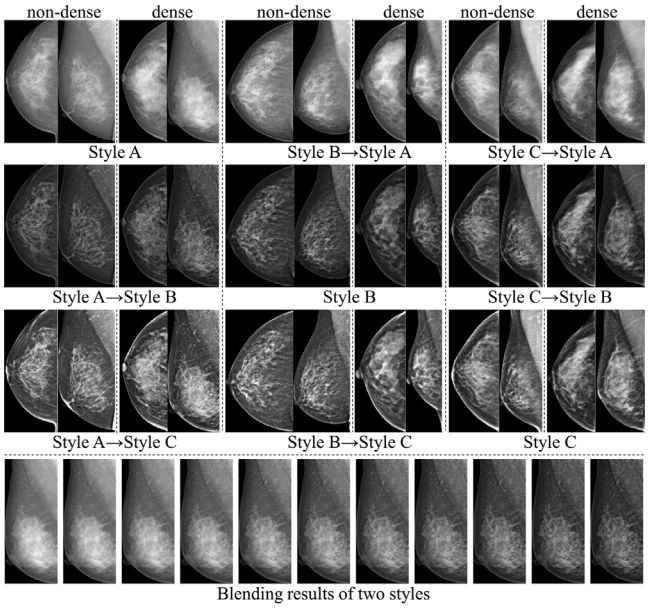


**Fig. 4.** Visual comparison of style transfer results. The first three rows display the results of style transfer using original images from style A, B, and C. The left column displays generated outcomes of A→B and A→C, whereas the center column represents generated outcomes of B→A and B→C. The right column presents the generated results of C→A and C→B. Each style contains both non-dense and dense breast cases. The last row demonstrates the series of synthesized styles from styles A and B with blending method shown in Eq. (2).

### 3.2. Multi-style and multi-view contrastive learning

To seek feature embedding with better generalizability for the vendor-style and view domains, MSVCL+ is devised. The synergy of the two learning schemes is illustrated in Fig. 3. Specifically, CycleGAN is employed to diversify the vendor styles of the original data. The style diversification is further augmented with an image blending operator. Afterward, the positive and negative pairs are packaged for contrastive

learning in the multi-style and multi-view schemes, respectively. The synergized contrastive learning scheme is driven by the optimization of both multi-style and multi-view contrastive losses.

#### 3.2.1. Multi-style contrastive learning

The image styles vary for different vendors, see Fig. 1. Therefore, the various vendor styles are regarded as distinctive domains. To endow the backbones of downstream tasks with better generalizability, we exploit the CycleGAN [27] augmented with a blending technique as the diversifying operation for contrastive learning. Specifically, given $M$ seen vendor-style domains, we train $\binom{M}{2}$ generators, which map the data distribution of source domain $\Omega_i$ to target $\Omega_j$ domain, $\forall i, j \in M$. With the $\binom{M}{2}$ generators, each of the original $N$ images in the training set can be diversified into $M-1$ transferred images, which is illustrated in the first three rows of Fig. 4. Moreover, we further use the image blending technique to achieve a smooth transition between two styles:

$$style_{blend} = style_A \times (1.0 - alpha) + style_B \times alpha, \quad (2)$$

where $alpha$ is the interpolation factor randomly selected from a set of values ranging from 0 to 1 with an interval of 0.1, $style_A$ corresponds to the style representation of an image from domain A, while $style_B$ represents the style from domain B. The blending results are illustrated in the last row of Fig. 4. In practice, for each original image, the samples for the bundling of positive and negative pairs are randomly picked from the pool of $L = \binom{M}{2} \times 9 + M$ blended style. With the aid of this blending technique, the diversity of the style samples for the MSCL+ will be significantly enhanced. Compared to the previous MSCL, the number of style samples has increased by 10 times.

In MSCL+, the two styles of the same source image (e.g., style A and style B from the CC view of a patient) are attracted together, see stage (a) in Fig. 2, while the same view of different patients (e.g., CCs of two patients, regardless of styles) are repelled from each other. The positive pairs for contrastive learning are constituted with any two images diversified from the same source image in the original $N$ image set. Therefore, there are possible $N \times \binom{L}{2}$ positive pairs available for random selection in contrastive learning. Referring to the right box in Fig. 3, the bundling strategy for negative pairs is further refined by only considering the same view positions from different patients, e.g., CC views from patient 1 and patient 2, to exclude the factor of domain gap between distinct views. The combination of CC and MLO views from different patients for a negative pair, which were involved in previous MSCL, may confuse contrastive learning and not be informative samples for effective training. The contrastive learning is then carried out by minimizing the Eq. (1) to seek feature embedding space with better generalizability to various vendor domains.

#### 3.2.2. Multi-view contrastive learning

A standard examination of mammography consists of CC and MLO views for each breast. Because the two standard views are taken from different angles of the same breast, they are mutually complementary for diagnosis. To seek domain-invariant feature embedding against different view domains, we explore the contrastive learning scheme to consider the distinctive view domains. Specifically, the CC and MLO views of the same breast from the same patient (e.g., LCC and LMLO of a patient) are treated as a positive pair, whereas the other combination of the CC and MLO views from different patients (e.g., LCC of patient 1 and LMLO of patient 2) is a negative pair. Similarly to MSCL+, the bundling strategy for negative pairs is further restricted to exclude the pairing of the same view positions from different patients, e.g., LCC views of patients 1 and 2, in MVCL+. This type of pairing was adopted in previous MVCL and may be redundant for training. To further enrich the sample diversity, we implement style-diversifying operations for the CC and MLO in each positive or negative pair. With the prepared sample pairs, MVCL+ can be carried out for the embedding of view-invariant features. The outperformance of MSCL+ and MVCL+ to the previous versions will be shown in the ablation experiments.

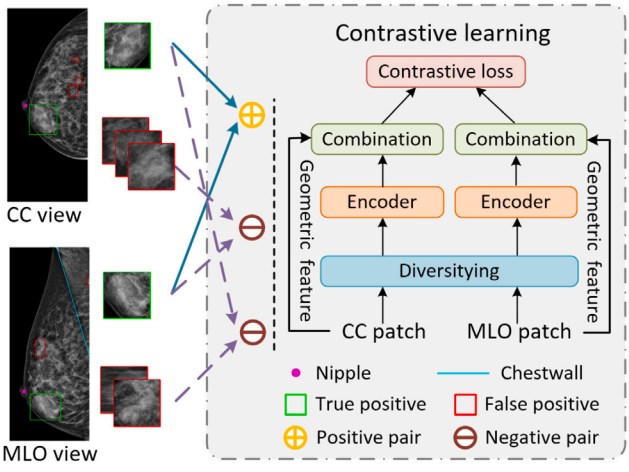


**Fig. 5.** The multi-view mass matching scheme with contrastive learning. The green boxes indicate regions of true positive mass, while the red boxes refer to false positives. In addition, the purple dots represent the location of the nipple, and the blue line suggests the chestwall.

### 3.3. Downstream tasks

#### 3.3.1. Mass detection

In this study, the classic single-view detection network of FCOS [53] is employed to identify mass in mammography. The derived pre-trained model from the self-supervised learning stage is adopted as the backbone of the FCOS architecture for the mass detection tasks. The detected results from CC and MLO paired images are two sets of candidate bounding boxes: $B_{cc} = \{b^1_{cc}, \ldots, b^i_{cc}, \ldots, b^N_{cc}\}$ and $B_{mlo} = \{b^1_{mlo}, \ldots, b^j_{mlo}, \ldots, b^M_{mlo}\}$.

#### 3.3.2. Mass matching

In the clinical reading, two found mass instances in the CC and MLO views are regarded as the same mass if several anatomical metrics, e.g., distance to nipple, appearance, and shape, are satisfied. If a mass can be found in both views, the diagnostic confidence is higher. On the other hand, the contrastive learning technique naturally takes two inputs for pair matching and can potentially be applied to realize the mass matching process. Inspired by this, we address mass matching with the incorporation of anatomical cues in contrastive learning. Specifically, the two sets of true-positive (TP) and false-positive (FP) bounding boxes from the FCOS are defined based on the IoU metrics with the ground truth, and these boxes can be further named as $TP_{cc}$, $FP_{cc}$, $TP_{mlo}$ and $FP_{mlo}$ w.r.t. MLO and CC views. The pair of $TP_{cc}$ and $TP_{mlo}$ that satisfies anatomical metrics is treated as a positive pair with the matching label Y of 1. On the other hand, the combination of boxes with any one from either $FP_{cc}$ or $FP_{mlo}$, is regarded as a negative pair with label $Y$ of 0.

As shown in Fig. 5, the CC/MLO patch matching is based on the metrics of appearance similarity, distances to the nipple and chestwall, and the box size. In the training process, the max margin contrastive loss is adopted as the matching error:

$$L_{mat} = \frac{1}{2K} \sum_{k=1}^{K} Y D^2 + (1 - Y) max(m - D, 0)^2, \quad (3)$$

where $D = \|f_1 - f_2\|_2$ refers to the L2 distance between the embedding features of paired samples. The label $Y$ is 1 if the two samples are matched, otherwise 0. The margin $m$ is a hyper-parameter, which suggests the lower bound distance between patches that are not matched.

**Table 1**
Distributions of data usage for different stages.

| Domain | Dataset | Vendor | Style transfer | Self-supervision | Downstream tasks (train/val/test) |
|---|---|---|---|---|---|
| Seen | Style A | GE | 1000 | 8000 | 600/100/100 |
| | Style B | UIH | 1000 | 8000 | 600/100/100 |
| | Style C | HOLOGIC | 1000 | 8000 | 600/100/100 |
| Unseen | Style D | SIEMENS | 0 | 0 | 0/0/100 |
| | Style E | INbreast | 0 | 0 | 0/0/100 |
| | Style F | DDSM | 0 | 0 | 0/0/100 |

**Table 2**
Ablation analysis w.r.t. various pre-training settings for mass detection tasks. The assessment metric is mAP [%].

| Pre-training method | Pre-training dataset | Seen domain | | | | Unseen domain | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Style A | Style B | Style C | Avg. | Style D | Style E | Style F | Avg. |
| Random | None | 59.8 | 70.5 | 66.5 | 65.6 ± 5.4 | 51.3 | 79.9 | 65.8 | 65.7 ± 14.3 |
| Supervised | ImageNet | 84.0 | 85.7 | 79.1 | 82.9 ± 3.4 | 82.4 | 76.3 | 81.0 | 79.9 ± 3.2 |
| SimCLR | Mammo | 84.4 | 91.6 | 84.1 | 86.7 ± 4.2 | 82.2 | 77.3 | 75.0 | 78.2 ± 3.7 |
| SimCLR | ImageNet → Mammo | 86.9 | 92.0 | 86.0 | 88.3 ± 3.2 | 86.2 | 80.5 | 84.2 | 83.6 ± 2.9 |
| MSCL | ImageNet → Mammo | 87.6 | 92.3 | 85.8 | 88.6 ± 3.4 | 88.5 | 85.2 | 86.0 | 86.6 ± 1.7 |
| MVCL | ImageNet → Mammo | 88.9 | 92.4 | 84.3 | 88.5 ± 4.1 | 86.7 | 85.3 | 87.3 | 86.4 ± 1.0 |
| MSVCL | ImageNet → Mammo | 91.8 | 94.0 | **89.1** | **91.6 ± 2.5** | 87.3 | **89.7** | 88.4 | 88.5 ± 1.2 |
| MSCL+ | ImageNet → Mammo | 92.2 | **94.3** | 86.7 | 91.1 ± 3.9 | 89.5 | 86.2 | 89.0 | 88.2 ± 1.8 |
| MVCL+ | ImageNet → Mammo | 90.8 | 94.1 | 86.7 | 90.5 ± 3.7 | 89.7 | 86.7 | 90.4 | 88.9 ± 2.0 |
| MSVCL+ | ImageNet → Mammo | **92.9** | 93.8 | 87.0 | 91.2 ± 3.7 | **90.4** | 89.4 | **94.1** | **91.3 ± 2.5** |

**Table 3**
Ablation analysis w.r.t. various pre-training settings for the four tasks. The assessment metrics are mAP [%] for mass detection, and acc [%] for the tasks of mass matching, BI-RADS rating, and breast density classification.

| Pre-training method | Pre-training dataset | Seen domain | | | | Unseen domain | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Detection | Match | BI-RADS | Density | Detection | Match | BI-RADS | Density |
| Random | None | 65.6 ± 5.4 | 75.0 ± 1.4 | 55.1 ± 3.9 | 65.3 ± 1.2 | 65.7 ± 14.3 | 60.3 ± 4.7 | 66.1 ± 34.7 | 40.3 ± 12.6 |
| Supervised | ImageNet | 82.9 ± 3.4 | 86.3 ± 5.8 | 85.4 ± 3.5 | 90.3 ± 2.1 | 79.9 ± 3.2 | 87.4 ± 3.4 | 81.4 ± 17.1 | 77.0 ± 6.6 |
| SimCLR | Mammo | 86.7 ± 4.2 | 79.7 ± 3.7 | 79.4 ± 4.7 | 92.3 ± 1.2 | 78.2 ± 3.7 | 81.5 ± 7.6 | 79.5 ± 17.2 | 73.0 ± 19.2 |
| SimCLR | ImageNet → Mammo | 88.3 ± 3.2 | 80.0 ± 6.7 | 87.4 ± 6.7 | 91.7 ± 2.5 | 83.6 ± 2.9 | 85.0 ± 7.4 | 83.3 ± 18.4 | 76.3 ± 16.3 |
| MSCL | ImageNet → Mammo | 88.6 ± 3.4 | 85.7 ± 4.1 | 86.4 ± 3.1 | 91.7 ± 1.5 | 86.6 ± 1.7 | 87.1 ± 6.6 | 85.8 ± 13.5 | 80.3 ± 6.2 |
| MVCL | ImageNet → Mammo | 88.5 ± 4.1 | 89.3 ± 2.9 | 89.2 ± 4.8 | 90.7 ± 0.6 | 86.4 ± 1.0 | 84.6 ± 7.1 | 84.2 ± 18.3 | 80.3 ± 6.6 |
| MSVCL | ImageNet → Mammo | **91.6 ± 2.5** | 89.6 ± 4.3 | 89.3 ± 4.9 | 91.7 ± 3.2 | 88.5 ± 1.2 | 91.5 ± 2.9 | 85.9 ± 9.4 | 81.0 ± 4.4 |
| MSCL+ | ImageNet → Mammo | 91.1 ± 3.9 | **92.2 ± 1.9** | **90.0 ± 4.5** | **93.0 ± 1.0** | 88.2 ± 1.8 | 89.7 ± 4.2 | 86.8 ± 9.4 | 82.3 ± 5.8 |
| MVCL+ | ImageNet → Mammo | 90.5 ± 3.7 | 89.0 ± 7.0 | 87.6 ± 3.5 | 92.3 ± 1.5 | 88.9 ± 2.0 | 86.9 ± 5.9 | 86.7 ± 14.8 | 83.3 ± 5.6 |
| MSVCL+ | ImageNet → Mammo | 91.2 ± 3.7 | 90.6 ± 3.1 | 89.0 ± 5.5 | 92.0 ± 2.6 | **91.3 ± 2.5** | **92.9 ± 2.5** | **87.8 ± 9.3** | **84.3 ± 8.3** |

#### 3.3.3. BI-RADS rating and breast density classification

In addition to mass detection and matching, we also illustrate the domain generalization performance on the two major downstream classification tasks of mammography analysis, i.e., BI-RADS rating and breast density classification. For BI-RADS rating, we adopt the pre-trained model for the classification model with five BI-RADS scores of 2∼3, 4 A, 4B, 4C, and 5. In the context of breast density classification, the pre-trained model is also adopted for the classification model with the input of a whole mammogram, either CC or MLO view. The breast density classes are non-dense and dense, i.e., the density categories of A∼B and C∼D.

## 4. Experiments and results

### 4.1. Datasets

Table 1 shows the details of data usage in this paper. One in-house and two public datasets are involved for model training and performance evaluation. The in-house dataset was collected from machines of four vendors: GE, United Imaging Healthcare (UIH), Hologic, and Siemens, denoted as A, B, C, and D, respectively, for short. All image data from four vendors were acquired from Asian women. The data from vendors A, B, and C are set as seen domains, whereas vendor D is treated as an unseen domain. The annotation for the in-house dataset was first carried out by two radiologists with 3–5 years of experience and then reviewed by a senior radiologist with more than 10 years of experience. To further evaluate the generalizability of the proposed method, the two public datasets, i.e., INbreast [54] and DDSM [55], denoted as E and F, respectively, are treated as the unseen datasets. In total, 29,700 mammograms (14,850 CC/MLO pairs) are involved in this study. 27,000 unannotated images are used for the training of stage (a) with style transfer and contrastive learning schemes. Specifically, all of the unannotated images are from the seen domain, including styles A, B, and C. The number of images used in each style is 1000 for style transfer and 8000 for self-supervision, respectively. The remaining 2700 annotated images are employed for the training, validation, and testing of stage (b), i.e., the downstream tasks. For each style in the seen domain, there are 600, 100, and 100 images for training, validation, and testing, respectively. For each style in the unseen domain, there are 100 images for testing. Datasets A, B, C, D, and E are tested on four downstream tasks, while dataset F is tested on three downstream tasks without BI-RADS rating, because the BI-RADS scores are not available in the DDSM.

### 4.2. Implementation details

In stage (a), the generator of the CycleGAN is a ResNet with 9 blocks of 20 convolution layers, whereas the discriminator is PatchGAN composed of 6 convolutional layers. The loss functions for the generator and discriminator are L1 and MSE, respectively. Random cropping of image patches with the size of $512 \times 512$ is implemented in the training process. For data standardization, a preprocessing step is performed

**Table 4**
Performance comparison among MSVCL+ and other SOTA domain generalization methods w.r.t mass detection task. The assessment metric is mAP [%].

| Method | Seen domain | | | | Unseen domain | | | |
|---|---|---|---|---|---|---|---|---|
| | Style A | Style B | Style C | Avg. | Style D | Style E | Style F | Avg. |
| Baseline | 86.9 | 92.0 | 86.0 | 88.3 ± 3.2 | 86.2 | 80.5 | 84.2 | 83.6 ± 2.9 |
| BigAug | 88.8 | 89.9 | 84.7 | 87.8 ± 2.7 | 88.4 | 86.5 | 84.3 | 86.4 ± 2.1 |
| DD | 86.9 | 90.7 | 86.1 | 87.9 ± 2.5 | 87.4 | 87.9 | 85.8 | 87.0 ± 1.1 |
| EISNet | 89.3 | 89.1 | 82.8 | 87.1 ± 3.7 | 86.2 | 83.8 | 85.6 | 85.2 ± 1.2 |
| MSVCL | 91.8 | **94.0** | **89.1** | **91.6 ± 2.5** | 87.3 | **89.7** | 88.4 | 88.5 ± 1.2 |
| MSVCL+(ours) | **92.9** | 93.8 | 87.0 | 91.2 ± 3.7 | **90.4** | 89.4 | **94.1** | **91.3 ± 2.5** |

**Table 5**
Performance comparison among MSVCL+ and other SOTA domain generalization methods w.r.t. the four tasks. The assessment metrics for the four tasks are the same with Table 3.

| Method | Seen domain | | | | Unseen domain | | | |
|---|---|---|---|---|---|---|---|---|
| | Detection | Match | BI-RADS | Density | Detection | Match | BI-RADS | Density |
| Baseline | 88.3 ± 3.2 | 80.0 ± 6.7 | 87.4 ± 6.7 | 91.7 ± 2.5 | 83.6 ± 2.9 | 85.0 ± 7.4 | 83.3 ± 18.4 | 76.3 ± 16.3 |
| BigAug | 87.8 ± 2.7 | 88.9 ± 3.2 | 89.0 ± 4.4 | 90.0 ± 4.0 | 86.4 ± 2.1 | 87.9 ± 6.6 | 83.3 ± 17.0 | 78.7 ± 5.0 |
| DD | 87.9 ± 2.5 | 84.4 ± 4.8 | 87.5 ± 8.3 | 90.3 ± 5.5 | 87.0 ± 1.1 | 88.5 ± 2.7 | 85.3 ± 14.2 | 82.0 ± 6.2 |
| EISNet | 87.1 ± 3.7 | 86.8 ± 2.7 | **89.6 ± 4.7** | 88.7 ± 7.1 | 85.2 ± 1.2 | 90.9 ± 2.8 | 82.4 ± 17.0 | 80.7 ± 7.6 |
| MSVCL | **91.6 ± 2.5** | 89.6 ± 4.3 | 89.3 ± 4.9 | 91.7 ± 3.2 | 88.5 ± 1.2 | 91.5 ± 2.9 | 85.9 ± 9.4 | 81.0 ± 4.4 |
| MSVCL+(ours) | 91.2 ± 3.7 | **90.6 ± 3.1** | 89.0 ± 5.5 | **92.0 ± 2.6** | **91.3 ± 2.5** | **92.9 ± 2.5** | **87.8 ± 9.3** | **84.3 ± 8.3** |

to align all various mammograms into the physical pixel spacing of 0.1 mm. The epoch is empirically set to 100 for the training of CycleGAN, which can lead to the best mass detection performance in the validation dataset. For balanced training, each pass of style transfer, e.g., style A to style B, the training data of the source and target styles are set to the same number. Codes and models of style transfer are available at: https://github.com/lizheren/MSVCL_PLUS.

The backbone model for the contrastive learning scheme and downstream tasks is ResNet-50. For fair comparison, the learning rate and batch size for all practices of the contrastive learning scheme are set at 0.3 and 256, respectively. Meanwhile, all contrastive learnings in all experiments share the same diversifying operations, including random cropping, random rotation in $\pm 10°$, horizontal flipping, and random color jittering (strength = 0.2).

The model backbones of all downstream tasks are initialized with the pre-trained models from the self-learning scheme, i.e., stage (a) in Fig. 2. For the training of FCOS models, the SGD optimization method is adopted with the parameters of learning rate, weight decay, and momentum set as 0.005, $10^4$, and 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 and scaling, are also implemented in the training of FCOS.

For the training of the remaining downstream tasks, the SGD method is also employed, with the parameters of learning rate, weight decay, and momentum set as 0.001, $10^5$, and 0.9, respectively. The epoch and batch size are set to 50 and 128, respectively. For the tasks of mass matching and BI-RADS rating, the input of the model is the squared ROI derived from the box identified by the detection network. Specifically, the squared ROI is first defined by taking the largest side length of the detected box with the same center. To consider the context, the squared ROI is further enlarged by 20%. For the task of breast density classification, the input to the model is the original mammograms. For these three tasks, the input images are resized to $224 \times 224$ pixels with the data augmentation of random flipping and scaling in the training process. The contrastive learning model for the mass matching task is trained with the setting of the hyper-parameter margin $m$ as 10 in Eq. (2). To facilitate the downstream mass matching task, a nipple and muscle segmentation model is employed for the computation of object-to-nipple and object-to-chestwall distances.

### 4.3. Ablation study

Table 2 reports the ablation study results on mass detection tasks w.r.t. different styles, and Table 3 shows the ablation study results of

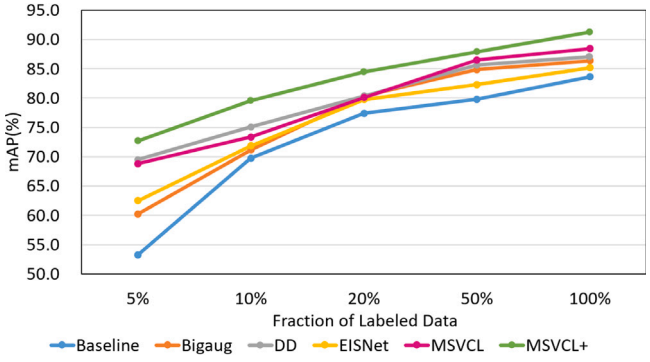


**Fig. 6.** Performance comparison for all implemented domain generalization methods in data hungry experiments.

averaged performances of the four downstream tasks over the seen and unseen domains. The ablation experiments are considering three major aspects: (1) pre-trained models, (2) various combinative contrastive learning schemes on style and view domains, and (3) revised versions of contrastive learning compared to our previous work [15]. Specifically, the performance comparison of different pre-trained models can be found in the first four rows of the Tables 2 and 3. The first row stands for training from scratch, whereas the second row indicates performances with a pre-trained model from ImageNet. The third "Mammo" row suggests that the SimCLR is trained from scratch with the unlabeled images of vendors A, B, and C. The pre-trained model of the fourth row, "ImageNet → Mammo" is derived by the SimCLR, which was initialized with ImageNet parameters and tuned with the Mammo set used in the third rows. As can be seen, the pre-trained model from "ImageNet → Mammo" derived by SimCLR is helpful for mass detection and other tasks.

By observing the groups of 5th to 7th rows and 8th to 10th rows in Tables 2 and 3, the performance can be mostly boosted by considering both style and view domains in the contrastive learning. Meanwhile, it may also be found that the pre-trained models from the revised versions, i.e., MSCL+, MVCL+, and MSVCL+, can mostly attain better performance on the five downstream tasks, particularly on unseen domains. The underlying reasons for the performance boost may be twofold. First, the increase in style diversity by the blending method may be helpful in seeking better embedding of style-invariant

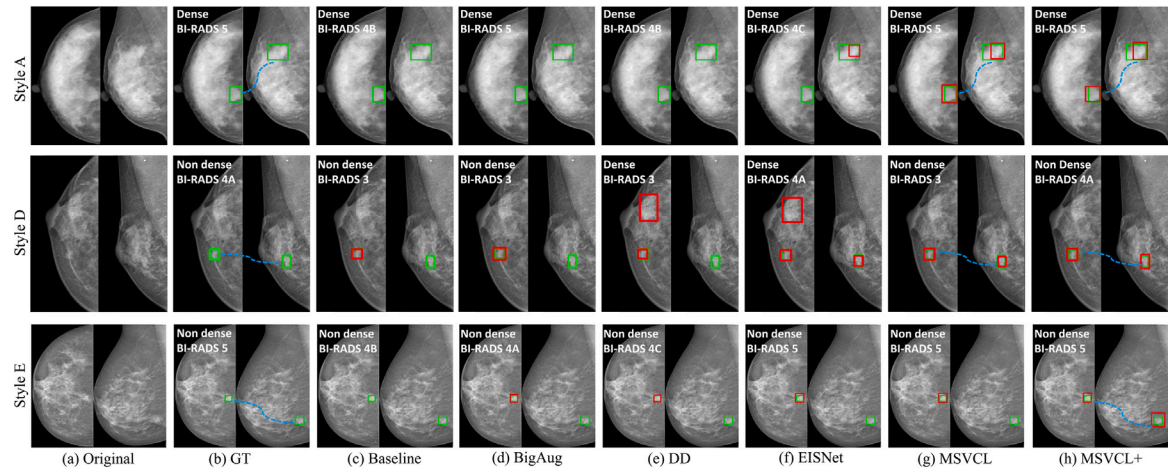


**Fig. 7.** Visual comparison of downstream tasks results in different styles w.r.t. various implemented methods. The green boxes refer to the mass regions labeled by radiologists, while the red boxes indicate the mass regions detected by computerized methods. The blue dotted lines indicate the outcomes of mass matching. The BI-RADS rating and breast density classification results are annotated at the top of the CC views.

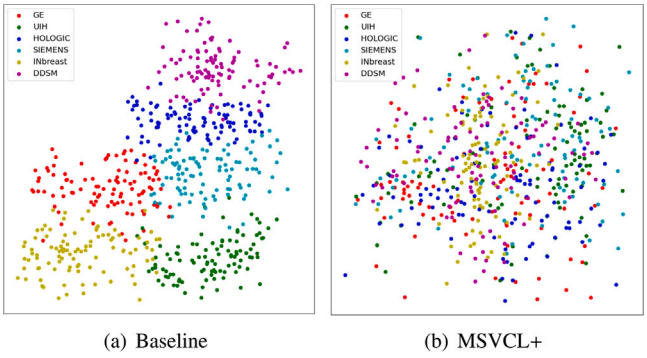


**Fig. 8.** t-SNE visualization for the pre-trained models from Baseline and MSVCL+.

features. Second, the new selection strategy provides more reasonable and informative negative pairs for contrastive learning that may render the sought feature embedding a better representative for the mammographic image domain.

### 4.4. Comparison with state-of-the-art (SOTA) methods

To further compare with other domain generalization methods, three SOTA methods, i.e., BigAug [28], Domain Diversification (DD) [19], and EISNet [21] are implemented. The BigAug [28] is a conventional data augmentation method, whereas the DD [19] proposes a generative learning method for domain generalization. The EISNet [21] 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. For a fair comparison, the other methods, including BigAug, DD, and EISNet, were fine-tuned with the pre-trained backbone of SimCLR on "ImageNet → Mammo", to ensure the use of the same amount of data. Table 4 reports the comparison results on the mass detection task with different styles, whereas Table 5 presents the performance comparison for the four downstream tasks across both seen and unseen domains. The Baseline rows in Tables 4 and 5 suggest the results of SimCLR with "ImageNet → Mammo". As can be found in Table 5, the MSVCL+ method achieves the best performance on unseen domains for the four downstream tasks.

A data-hungry experiment is also conducted for the comparison of SOTA methods. Five fraction settings, i.e., 5%, 10%, 20%, 50%, and 100% of labeled (training) data, are applied for the experiment to illustrate the learning capability of domain generalization methods in terms of the amount of training data. The data-hungry experiment is conducted on the detection task, and the corresponding results are shown in Fig. 6. As can be observed, the MSVCL+ outperforms other methods even with a very small amount of data, i.e., the 5% fraction setting. Fig. 7 lists several results of the implemented downstream tasks from different domain generalization methods for visual comparison. It can be observed that MSVCL+ can help the tasks attain better performance.

### 4.5. Visualization of feature space

The t-SNE [56] is employed to illustrate the feature distribution of various vendor domains for visual evaluation of generalizability. In Fig. 8, the t-SNE plots compare the distributions of Baseline and MSVCL+, where features of various vendors are masked with distinctive colors. As can be seen, the vendor features of MSVCL+ are well mixed. It may be suggested that the generalization w.r.t. the vendor domain is relatively promising.

## 5. Discussion and conclusion

The effectiveness of our style-based augmentations in improving contrastive learning can be understood through the lens of representation learning theory. First, the Style-based augmentations encourage the model to learn representations that are invariant to vendor-specific image characteristics while remaining equivariant to clinically relevant features. This aligns with the principle that good representations should be sensitive to important variations in the input while being robust to nuisance factors (in this case, vendor-specific imaging characteristics). For example, the diagnosis information should be vendor invariant. Moreover, contrastive learning can be viewed as maximizing mutual information between different views of the data, which is successfully demonstrated in previous works. By introducing style variations, we create views that share high-level semantic content but differ in low-level statistics. This challenges the model to extract more abstract, semantically meaningful features, potentially leading to more robust representations.

A novel domain generalization method, denoted as MSVCL+, is proposed to endow the deep learning models with better generalizability to the vendor style and view domains. The MSVCL+ has been shown to be helpful for four mammographic image analysis tasks, i.e., detection, matching, BI-RADS rating, and breast density classification. The MSVCL+ can provide more robust feature embedding against various

vendor-style domains. Experimental results suggest that our method can effectively improve mammographic image analysis tasks on unseen domains compared to the four implemented SOTA domain generalization methods. In particular, our method achieves the best performance on the public datasets INbreast and DDSM, where the domain gaps may not only include image styles but also population factors. The INbreast and DDSM datasets were collected from Western women. In contrast, our methods were trained on mammograms acquired from Asian women. Therefore, the generalization efficacy of our proposed method is further corroborated.

## CRediT authorship contribution statement

**Zheren Li:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Zhiming Cui:** Methodology, Conceptualization. **Lichi Zhang:** Writing – review & editing, Supervision, Conceptualization. **Sheng Wang:** Visualization, Data curation. **Chenjin Lei:** Validation, Data curation. **Xi Ouyang:** Writing – review & editing, Conceptualization. **Dongdong Chen:** Investigation. **Xiangyu Zhao:** Investigation. **Chunling Liu:** Resources. **Zaiyi Liu:** Resources. **Yajia Gu:** Resources. **Dinggang Shen:** Writing – review & editing, Funding acquisition, Conceptualization. **Jie-Zhi Cheng:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work was supported in part by the Key Research and Development Program of Guangdong Province, China, under Grant 2021B0101420006; in part by Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application under Grant 2022B1212010011; in part by the National Natural Science Foundation of China under Grant 82171920; and in part by the National Natural Science Foundation of China under Grant 82071878.

## References

[1] R.L. Siegel, K.D. Miller, H.E. Fuchs, A. Jemal, Cancer statistics, 2022, CA 72 (1) (2022) 7–33.

[2] W. Lotter, A.R. Diab, B. Haslam, J.G. Kim, G. Grisot, E. Wu, K. Wu, J.O. Onieva, Y. Boyer, J.L. Boxerman, et al., Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach, Nature Med. (2021) 1–6.

[3] S.M. McKinney, M. Sieniek, V. Godbole, J. Godwin, N. Antropova, H. Ashrafian, T. Back, M. Chesus, G.C. Corrado, A. Darzi, et al., International evaluation of an AI system for breast cancer screening, Nature 577 (7788) (2020) 89–94.

[4] M. Salim, E. Wåhlin, K. Dembrower, E. Azavedo, T. Foukakis, Y. Liu, K. Smith, M. Eklund, F. Strand, External evaluation of 3 commercial artificial intelligence algorithms for independent assessment of screening mammograms, JAMA Oncol. 6 (10) (2020) 1581–1588.

[5] X. Ouyang, J. Che, Q. Chen, Z. Li, Y. Zhan, Z. Xue, Q. Wang, J.-Z. Cheng, D. Shen, Self-adversarial learning for detection of clustered microcalcifications in mammograms, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VII 24, Springer, 2021, pp. 78–87.

[6] L. Abdelrahman, M. Al Ghamdi, F. Collado-Mesa, M. Abdel-Mottaleb, Convolutional neural networks for breast cancer detection in mammography: A survey, Comput. Biol. Med. 131 (2021) 104248.

[7] Y. Yan, P.-H. Conze, M. Lamard, G. Quellec, B. Cochener, G. Coatrieux, Towards improved breast mass detection using dual-view mammogram matching, Med. Image Anal. 71 (2021) 102083.

[8] Q. Wu, H. Tan, Y. Wu, P. Dong, J. Che, Z. Li, C. Lei, D. Shen, Z. Xue, M. Wang, Whole mammography diagnosis via multi-instance supervised discriminative localization and classification, in: Machine Learning in Medical Imaging: 13th International Workshop, MLMI 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18, 2022, Proceedings, Springer, 2022, pp. 131–139.

[9] W. Zhao, R. Wang, Y. Qi, M. Lou, Y. Wang, Y. Yang, X. Deng, Y. Ma, Bascnet: Bilateral adaptive spatial and channel attention network for breast density classification in the mammogram, Biomed. Signal Process. Control 70 (2021) 103073, http://dx.doi.org/10.1016/j.bspc.2021.103073, URL: https://www.sciencedirect.com/science/article/pii/S1746809421006704.

[10] J. Xing, Z. Li, B. Wang, Y. Qi, B. Yu, F.G. Zanjani, A. Zheng, R. Duits, T. Tan, Lesion segmentation in ultrasound using semi-pixel-wise cycle generative adversarial nets, IEEE/ACM Trans. Comput. Biol. Bioinform. (2020).

[11] L. Garrucho, K. Kushibar, S. Jouide, O. Diaz, L. Igual, K. Lekadir, Domain generalization in deep learning based mass detection in mammography: A large-scale multi-center study, Artif. Intell. Med. 132 (2022) 102386.

[12] X. Liu, B. Glocker, M.M. McCradden, M. Ghassemi, A.K. Denniston, L. Oakden-Rayner, The medical algorithmic audit, Lancet Digit. Health (2022).

[13] D. Krueger, E. Caballero, J.-H. Jacobsen, A. Zhang, J. Binas, D. Zhang, R. Le Priol, A. Courville, Out-of-distribution generalization via risk extrapolation (rex), in: International Conference on Machine Learning, PMLR, 2021, pp. 5815–5826.

[14] X. Li, W. Zhang, H. Ma, Z. Luo, X. Li, Domain generalization in rotating machinery fault diagnostics using deep neural networks, Neurocomputing 403 (2020) 409–420.

[15] Z. Li, Z. Cui, S. Wang, Y. Qi, X. Ouyang, Q. Chen, Y. Yang, Z. Xue, D. Shen, J.-Z. Cheng, Domain generalization for mammography detection via multi-style and multi-view contrastive learning, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2021, pp. 98–108.

[16] E. Romera, L.M. Bergasa, J.M. Alvarez, M. Trivedi, Train here, deploy there: Robust segmentation in unseen domains, in: 2018 IEEE Intelligent Vehicles Symposium, IV, IEEE, 2018, pp. 1828–1833.

[17] R. Volpi, H. Namkoong, O. Sener, J.C. Duchi, V. Murino, S. Savarese, Generalizing to unseen domains via adversarial data augmentation, in: NeurIPS, 2018.

[18] X. Yue, Y. Zhang, S. Zhao, A. Sangiovanni-Vincentelli, K. Keutzer, B. Gong, 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, 2019, pp. 2100–2110.

[19] T. Kim, M. Jeong, S. Kim, S. Choi, C. Kim, 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, 2019, pp. 12456–12465.

[20] S. Zakharov, W. Kehl, S. Ilic, Deceptionnet: Network-driven domain randomization, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 532–541.

[21] S. Wang, L. Yu, C. Li, C.-W. Fu, P.-A. Heng, Learning from extrinsic and intrinsic supervisions for domain generalization, in: European Conference on Computer Vision, Springer, 2020, pp. 159–176.

[22] Q. Liu, Q. Dou, P.-A. Heng, Shape-aware meta-learning for generalizing prostate mri segmentation to unseen domains, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 475–485.

[23] Q. Dou, D. Coelho de Castro, K. Kamnitsas, B. Glocker, Domain generalization via model-agnostic learning of semantic features, Adv. Neural Inf. Process. Syst. 32 (2019).

[24] N. Chen, L. Liu, Z. Cui, R. Chen, D. Ceylan, C. Tu, W. Wang, Unsupervised learning of intrinsic structural representation points, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 9121–9130.

[25] S. Azizi, B. Mustafa, F. Ryan, Z. Beaver, J. Freyberg, J. Deaton, A. Loh, A. Karthikesalingam, S. Kornblith, T. Chen, et al., Big self-supervised models advance medical image classification, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 3478–3488.

[26] H. Sowrirajan, J. Yang, A.Y. Ng, P. Rajpurkar, Moco pretraining improves representation and transferability of chest x-ray models, in: Medical Imaging with Deep Learning, PMLR, 2021, pp. 728–744.

[27] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232.

[28] L. Zhang, X. Wang, D. Yang, T. Sanford, S. Harmon, B. Turkbey, B.J. Wood, H. Roth, A. Myronenko, D. Xu, et al., Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation, IEEE Trans. Med. Imaging 39 (7) (2020) 2531–2540.

[29] D. Mahajan, S. Tople, A. Sharma, Domain generalization using causal matching, in: International Conference on Machine Learning, PMLR, 2021, pp. 7313–7324.

[30] G. Blanchard, A.A. Deshmukh, U. Dogan, G. Lee, C. Scott, Domain generalization by marginal transfer learning, J. Mach. Learn. Res. (2021).

[31] S. Hu, K. Zhang, Z. Chen, L. Chan, Domain generalization via multidomain discriminant analysis, in: Uncertainty in Artificial Intelligence, PMLR, 2020, pp. 292–302.

[32] R. Gong, W. Li, Y. Chen, L.V. Gool, Dlow: Domain flow for adaptation and generalization, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 2477–2486.
[33] S. Zhao, M. Gong, T. Liu, H. Fu, D. Tao, Domain generalization via entropy regularization, Adv. Neural Inf. Process. Syst. 33 (2020) 16096–16107.
[34] S. Motiian, M. Piccirilli, D.A. Adjeroh, G. Doretto, Unified deep supervised domain adaptation and generalization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 5715–5725.
[35] X. Pan, P. Luo, J. Shi, X. Tang, Two at once: Enhancing learning and generalization capacities via ibn-net, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 464–479.
[36] X. Fan, Q. Wang, J. Ke, F. Yang, B. Gong, M. Zhou, Adversarially adaptive normalization for single domain generalization, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 8208–8217.
[37] K. Ahuja, E. Caballero, D. Zhang, J.-C. Gagnon-Audet, Y. Bengio, I. Mitliagkas, I. Rish, Invariance principle meets information bottleneck for out-of-distribution generalization, Adv. Neural Inf. Process. Syst. 34 (2021).
[38] M. Xu, L. Qin, W. Chen, S. Pu, L. Zhang, Multi-view adversarial discriminator: Mine the non-causal factors for object detection in unseen domains, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 8103–8112.
[39] R. Meng, X. Li, W. Chen, S. Yang, J. Song, X. Wang, L. Zhang, M. Song, D. Xie, S. Pu, Attention diversification for domain generalization, in: European Conference on Computer Vision, Springer, 2022, pp. 322–340.
[40] L. Zhang, L. Qin, M. Xu, W. Chen, S. Pu, W. Zhang, Randomized spectrum transformations for adapting object detector in unseen domains, IEEE Trans. Image Process. (2023).
[41] L. Jing, Y. Tian, Self-supervised visual feature learning with deep neural networks: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 43 (11) (2020) 4037–4058.
[42] F.M. Carlucci, A. D'Innocente, S. Bucci, B. Caputo, T. Tommasi, Domain generalization by solving jigsaw puzzles, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 2229–2238.
[43] D. Kim, Y. Yoo, S. Park, J. Kim, J. Lee, Selfreg: Self-supervised contrastive regularization for domain generalization, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 9619–9628.
[44] S. Jeon, K. Hong, P. Lee, J. Lee, H. Byun, Feature stylization and domain-aware contrastive learning for domain generalization, in: Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 22–31.
[45] T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A simple framework for contrastive learning of visual representations, in: International Conference on Machine Learning, PMLR, 2020, pp. 1597–1607.
[46] K. He, H. Fan, Y. Wu, S. Xie, R. Girshick, Momentum contrast for unsupervised visual representation learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 9729–9738.
[47] X. Chen, H. Fan, R. Girshick, K. He, Improved baselines with momentum contrastive learning, 2020, arXiv preprint arXiv:2003.04297.
[48] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar, et al., Bootstrap your own latent-a new approach to self-supervised learning, Adv. Neural Inf. Process. Syst. 33 (2020) 21271–21284.
[49] H. Jung, B. Kim, I. Lee, M. Yoo, J. Lee, S. Ham, O. Woo, J. Kang, Detection of masses in mammograms using a one-stage object detector based on a deep convolutional neural network, PLoS One 13 (9) (2018) e0203355.
[50] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2980–2988.
[51] R. Shen, J. Yao, K. Yan, K. Tian, C. Jiang, K. Zhou, Unsupervised domain adaptation with adversarial learning for mass detection in mammogram, Neurocomputing 393 (2020) 27–37.
[52] Z. Yang, Z. Cao, Y. Zhang, Y. Tang, X. Lin, R. Ouyang, M. Wu, M. Han, J. Xiao, L. Huang, et al., MommiNet-v2: Mammographic multi-view mass identification networks, Med. Image Anal. 73 (2021) 102204.
[53] Z. Tian, C. Shen, H. Chen, T. He, Fcos: A simple and strong anchor-free object detector, IEEE Trans. Pattern Anal. Mach. Intell. 44 (4) (2020) 1922–1933.
[54] I.C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M.J. Cardoso, J.S. Cardoso, Inbreast: toward a full-field digital mammographic database, Academic Radiol. 19 (2) (2012) 236–248.
[55] M. Heath, K. Bowyer, D. Kopans, R. Moore, W. Kegelmeyer, The digital database for screening mammography, in: Proceedings of the 5th international workshop on digital mammography, 2000, pp. 212–218.
[56] L. Van der Maaten, G. Hinton, Visualizing data using t-SNE., J. Mach. Learn. Res. 9 (11) (2008).