Towards Robust Dual-View Transformation via Densifying Sparse Supervision for Mammography Lesion Matching

Junlin Xian[1], Zhiwei Wang[2,3], Kwang-Ting Cheng[4], and Xin Yang[1,5(✉)]

[1] School of Electronic Information and Communication,
Huazhong University of Science and Technology, Wuhan, China
[2] Britton Chance Center for Biomedical Photonics, Wuhan National Laboratory
for Optoelectronics, Huazhong University of Science and Technology, Wuhan, China
[3] MoE Key Laboratory for Biomedical Photonics, Collaborative Innovation Center
for Biomedical Engineering, School of Engineering Sciences,
Huazhong University of Science and Technology, Wuhan, China
[4] Hong Kong University of Science and Technology, Hong Kong, China
[5] Wuhan National Laboratory for Optoelectronics, Huazhong University of Science
and Technology, Wuhan, China
`xinyang2014@hust.edu.cn`

Abstract. A holistic understanding of dual-view transformation (DVT) is an enabling technique for computer-aided diagnosis (CAD) of breast lesion in mammogram, e.g., micro-calcification (μC) or mass matching, dual-view feature extraction etc. Learning a complete DVT usually relies on a dense supervision which indicates a corresponding tissue in one view for each tissue in another. Since such dense supervision is infeasible to obtain in practical, a sparse supervision of some traceable lesion tissues across two views is thus an alternative but will lead to a defective DVT, limiting the performance of existing CAD systems dramatically. To address this problem, our solution is simple but very effective, i.e., densifying the existing sparse supervision by synthesizing lesions across two views. Specifically, a Gaussian model is first employed for capturing the spatial relationship of real lesions across two views, guiding a following proposed LT-GAN where to synthesize fake lesions. The proposed novel LT-GAN can not only synthesize visually realistic lesions, but also guarantee appearance consistency across views. At last, a denser supervision can be composed based on both real and synthetic lesions, enabling a robust DVT learning. Experimental results show that a DVT can be learned via our densified supervision, and thus result in a superior performance of cross-view μC matching on INbreast and CBIS-DDSM dataset to the state-of-the-art methods.

Keywords: Dual-view transformation · Supervision-densifying · Lesion matching

1   Introduction

Mammography consists of X-ray images usually from two views of breast, i.e., craniocaudal (CC) and mediolateral-oblique (MLO), which can provide important biomarkers for early breast cancer diagnosis. Holistically understanding dual-view transformation (DVT), which reflects how spatial location and appearance of each tissue changes across CC and MLO views, can act as a significant guidance for both manual mammography screening and computer-aided diagnosis (CAD) of breast cancer. For instance, radiologists often identify different lesions and false positives in one view by cross-checking with those in another.

However, learning a complete DVT must rely on a dense supervision which indicates every paired region in CC and MLO images projected from a same breast tissue. Such dense supervision is practically infeasible to obtain for information loss of the 3D construction during the mammogram [2]. Fortunately, a sparse supervision obtained by some visually traceable lesions, e.g., microcalcification (μC) or mass, has been demonstrated to be an alternative for learning DVT in recent works [12,17]. Shaked *et al.* [12] generated patch pairs for each mass lesion and used a CNN approach to learn pairwise feature representations. Yan *et al.* [17] further extended via an auxiliary classification task to tell if patches contain lesions or not.

The key of their success is Siamese network [5,7,15] whose mechanism of network sharing shows effectiveness in dual-view related tasks, e.g., cross-view matching. Depending on whether there is a metric network or not, the Siamese network-based approaches can be categorized into two classes. The former first utilizes a network to extract features from two regions and then predicts a probability that the two regions belong to the same object or not. For instance, MatchNet [5] used two weight-shared networks as a feature extractor and three fully-connected layers as a metric network. Zagoruyko *et al.* [20] concatenated two patches into a 2-channel meta-patch and fed it into a single-branch CNN as both feature extractor and metric network. SCFM [13] further exploited for combination in spatial-dimension. In contrast, the latter aims at reshaping the distribution of matched and unmatched samples, i.e., minimizing feature distances of matched samples and maximizing those of unmatched ones. Melekhov *et al.* [10] used Euclidean distance for pairwise distance measuring, while PN-Net [1] employed a triplet-loss.

Based on those Siamese network-based approaches, transformation between two views can be easily learned by exhaustively searching every matched region across different views. However, without a dense supervision, only a defective DVT can be learned via the Siamese network. For instance, previous works [9,12,17,19] can only learn an inadequate DVT around mass regions rather than a whole breast. To this end, we aim at synthesizing some trackable lesion tissues like μC or mass across two views. These synthesized cross-view lesions are guaranteed that 1) the correlation of spatial locations between two views is correct, and 2) changes of appearance across two views are consistent with those of real lesions under breast compression and view transferring. Both synthesized and real trackable lesion tissues make the original sparse supervision denser,
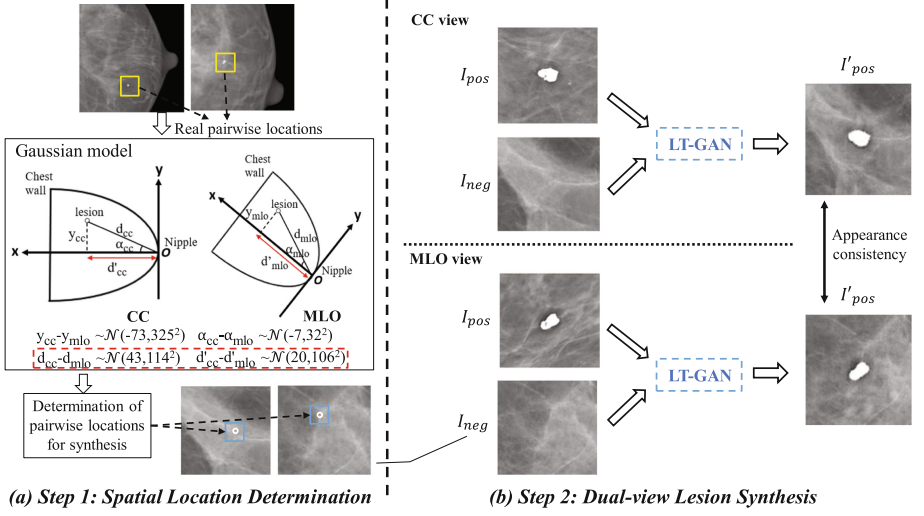
Fig. 1. An overall framework of our proposed supervision-densifying method. (a) illustrates the corresponding four Gaussian distributions in INbreast.

which thus leads to a more robust DVT learning. Specifically, our proposed framework for densifying sparse supervision is shown Fig. 1, which consists of two key steps. In the first step, a Gaussian-model is employed to analysis cross-view spatial correlation of real lesions, and further guide determination of candidate locations for synthesizing. Next, we propose a novel Lesion-Translation GAN (LT-GAN) based on Cycle-GAN [21] which can synthesize lesions by inheriting real pairwise lesion information so as to make cross-view changes of appearance realistic.

To summarize, major contributions of this work are as follows: (1) We introduce a novel supervision-densifying approach, which helps capture a robust DVT in mammogram. (2) We present a novel LT-GAN for realistic trackable lesion tissue synthesis, which prevents limitations of regular GANs and remains appearance consistency. (3) We conduct several experiments of cross-view μC matching on both INbreast [8] and CBIS-DDSM [11] dataset, demonstrating that our proposed supervision-densifying approach superior on various general architectures, and greatly surpasses state-of-the-art methods of mammography matching.

2   Method

Figure 1 outlines the overall framework of our proposed supervision-densifying approach, which consists of two key steps, i.e., a Gaussian model towards pairwise location determination, and a LT-GAN to synthesize realistic μC lesions across views. We will detail the two sequential steps in the following.

2.1   Spatial Location Determination in Fake Lesions

To fit a Gaussian model, we first build up an orthogonal coordinate system for a breast to parameterize spatial locations of lesions. We consider the nipple as the origin, the chest wall as the y-axis, and a line orthogonal to the chest wall as x-axis as shown in Fig. 1(a). Based on this coordinate system, we can have spatial parameters to describe the location of each lesion, e.g., d for pixel distance to the nipple, y (d') for projection distance of d on y-axis (x-axis) and α for the angle between d and d'. Instead of simply treating these parameters as constant across views [3,16], we assume the changes of parameters across views can be modeled by a Gaussian distribution, e.g., $d_{cc} - d_{mlo} \sim \mathcal{N}\left(\mu, \sigma^2\right)$. To estimate μ, σ^2 for each parameter, we collect real lesion pairs and obtain four Gaussian distributions. For instance, we have $d_{cc} - d_{mlo} \sim \mathcal{N}(43, 114^2)$, $d'_{cc} - d'_{mlo} \sim \mathcal{N}(20, 106^2)$, $y_{cc} - y_{mlo} \sim \mathcal{N}(-73, 325^2)$ and $\alpha_{cc} - \alpha_{mlo} \sim \mathcal{N}(-7, 32^2)$ for the INbreast dataset as shown in Fig. 1(a). The p-value of Kolmogorov-Smirnov test for each spatial parameter is 0.515, 0.94, 0.463 and 0.255 for d, d', y and α respectively, and larger p-value means higher degree of fitness to the corresponding Gaussian distribution. Therefore, we use d and d' with the largest two p-values to fully describe the spatial location of a lesion in INbreast. In the inference, we first randomly select a point (d_{mlo}, d'_{mlo}) in MLO view based on the orthogonal coordinate, and then generate the changes from MLO to CC view ($d_{cc} - d_{mlo}$, $d'_{cc} - d'_{mlo}$) using corresponding Gaussian models. By adding (d_{mlo}, d'_{mlo}) and ($d_{cc} - d_{mlo}$, $d'_{cc} - d'_{mlo}$) together, we can get a counterpart point in CC view (d_{cc}, d'_{cc}) of the original point in MLO view (d_{mlo}, d'_{mlo}). These paired locations can be used for guiding where to synthesize fake lesions.

2.2   LT-GAN for Cross-View Lesion Synthesis

As shown in Fig. 2, our proposed LT-GAN consists of two different domain-translation modules: a lesion-depositing translator (i.e., De-lesion) and a conditional generative translator (i.e., Conditional-GAN), denoted as T_{p2n} and T_{n2p} respectively. The two translators share identical backbone of U-net [14] but not trainable weights. Inspired by [18], T_{p2n} takes a patch cropped around the center of a lesion I_{pos} as input, and separate it as two parts, i.e., Z a 256-dimentional lesion-relative latent vectors and I'_{neg} a pure normal tissue of I_{pos}. Conversely, T_{n2p} recovers an encoded 256-d latent vector based on a pure background patch I_{neg} so as to generate a synthesized patch I'_{pos} with lesion in I_{pos} and normal tissue in I_{neg} combined together. As can be seen in Fig. 2, LT-GAN is trained in a weakly-supervised and cycling fashion, which only requires image-level labels indicating containing lesions or not instead of pixel-level annotations of each lesion. Loss functions designed for training LT-GAN are detailed below.

Adversarial Loss: For stable training based on unpaired I_{pos} and I_{neg}, we utilize two discriminators (D_{pos} and D_{neg}) and follow WGAN-GP [4] to measure W-distances between the fake and real samples in both domains:

$$W_{pos}\left(T_{n2p}, D_{pos}\right) = \sum_{x \in I_{pos}} \left[D_{pos}(\mathrm{x})\right] - \sum_{x' \in I'_{pos}} \left[D_{pos}\left(\mathrm{x}'\right)\right] - \mu_{pos} R_{pos} \quad (1)$$
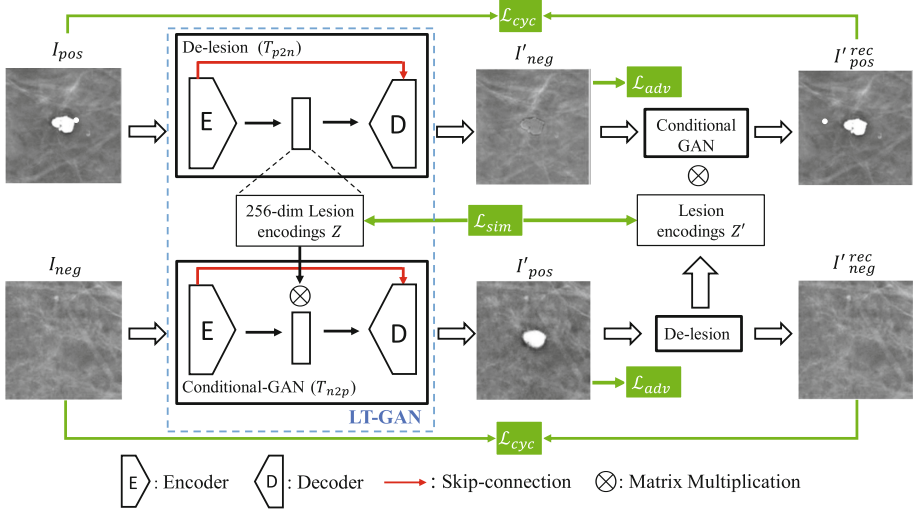
Fig. 2. An overview of our proposed LT-GAN training in a cycling fashion.

$$W_{neg}\left(T_{p2n}, D_{neg}\right) = \sum_{y \in I_{neg}} \left[D_{neg}(y)\right] - \sum_{y' \in I'_{neg}} \left[D_{neg}\left(y'\right)\right] - \mu_{neg} R_{neg} \quad (2)$$

where T_{n2p} (T_{p2n}) aims to minimize the above objective against an adversary D_{pos} (D_{neg}) that tries to maximize it, $\mu_{pos} R_{pos}$ $(\mu_{neg} R_{neg})$ is the gradient penalty term. Accordingly, the loss function of adversarial training translators is:

$$\mathcal{L}_{adv} = W_{pos} + W_{neg} \quad (3)$$

Cycle-Consistency Loss: Same as Cycle-GAN, we train our LT-GAN in a cycling-fashion so as to well preserve background consistency between fake and real images. Reconstruction patches of two domains can be denoted as:

$$I'^{\,rec}_{pos} = T_{n2p}\left(T_{p2n}\left(I_{pos}\right), Z'\right), I'^{\,rec}_{neg} = T_{p2n}\left(T_{n2p}\left(I_{neg}, Z\right)\right) \quad (4)$$

The overall cycle-consistency loss can be expressed as:

$$\mathcal{L}_{cyc}\left(T_{n2p}, T_{p2n}\right) = \left\|I'^{\,rec}_{pos} - I_{pos}\right\|_1 + \left\|I'^{\,rec}_{neg} - I_{neg}\right\|_1 \quad (5)$$

Lesion Similarity Loss: It is desirable that lesion-relative latent vectors Z and Z' should share same information, while existing noises during training usually bring inconsistency and oppositely confuse the synthesis process. To alleviate noise of lesion encodings extracted from T_{p2n}, we build up a L1-based similarity loss $\mathcal{L}_{sim}$ between lesion encodings Z and Z':

$$\mathcal{L}_{sim} = \|Z - Z'\|_1 \quad (6)$$

Other than above losses shown in Fig. 2, we additionally add two auxiliary losses: $\mathcal{L}_{app}$ and $\mathcal{L}_{idt}$. These two losses can effectively alleviate wrong mapping results and thus facilitate our training process.

Identity Loss: We follow Cycle-GAN to impose specific identity loss for each translator. For T_{p2n}, if given I_{neg} as input, its output should keep unchanged. Then for T_{n2p}, if lesion encodings $\hat{Z}$ are extracted from I_{neg} or there is no Z for fusion, its output should remain the same as input.

We formulate the total identity loss by measuring L1 distance between identity patches and inputs:

$$\mathcal{L}_{idt} = \left\| T_{n2p}\left(I_{neg}, \hat{Z}\right) - I_{neg} \right\|_1 + \left\| T_{n2p}\left(I_{neg}\right) - I_{neg} \right\|_1 + \left\| T_{p2n}\left(I_{neg}\right) - I_{neg} \right\|_1 \tag{7}$$

Appearance Translation Loss: To improve T_{n2p}'s tolerance of noise and force it to effectively fuse specific lesion information from Z, a L1-based appearance translation loss is applied for consistency between $I'_{pos} - I_{neg}$ and $I_{pos} - I'_{neg}$, where these two difference terms represent lesion apperance information from I'_{pos} and I_{pos} respectively:

$$\mathcal{L}_{app} = \left\| \left(I'_{pos} - I_{neg}\right) - \left(I_{pos} - I'_{neg}\right) \right\|_1 \tag{8}$$

Our goal of training is to minimize sum of above objectives:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{adv} + \lambda_2 \mathcal{L}_{cyc} + \lambda_3 \mathcal{L}_{idt} + \lambda_4 \mathcal{L}_{app} + \lambda_5 \mathcal{L}_{sim} \tag{9}$$

where λ_i (i=1,2,. . . ,5) are weights to control importance of different losses. We experimentally set λ_1=1, λ_2=1, λ_3=0.5, λ_4=40 and λ_5=1.

3   Experiment

3.1   Dataset and Implementation Details

To validate effectiveness of our proposed method on DVT learning, we conduct several experiments of cross-view μC matching and evaluate on two public mammography dataset, i.e., the INbreast (115 patients and 410 images in total) and the CBIS-DDSM (3103 images in total). We selected 38 (44) cases with a single lesion, and 9 (16) cases with multiple lesions in INbreast (CBIS-DDSM). Each case has paired CC and MLO view images. We preprocess the dataset by operating contrast enhancement and brightness normalization to reduce brightness variance across views.

For each ground-truth (GT) point of μC, we define a 200 * 200 square with the μC in the center as the GT mask. Candidate lesion patches are then generated by considering their Dice-Sørensen Coefficient (DSC) with the GT mask, i.e., true lesion patches if DSC > 0.25, and false lesion patches otherwise. For each

μC, we generate 7 true lesion patches and 3 false ones in each view. We pair these patches of two views, and obtain a positive pair which consists of true lesion patches of the same μC from two views, and a negative pair which is any other combination except two false lesion patches of two views.

Evaluations are based on 5-fold cross-validation and metrics are balanced-accuracy (i.e. b-Acc) and AUC for classifying positive and negative pairs. Each candidate patch is resized into 64 * 64 pixels. All matching networks are trained with the cross-entropy loss and optimized using Adam solver under a batch size of 64. The learning rate starts at 5e−5 and gradually decays by 0.9.

3.2   Ablation Study

First, we conduct an ablation study on INbreast to analysis impacts of two requirements of synthesis lesions, i.e., 1) the correlation of spatial locations between two views is correct, and 2) changes of appearance across two views are consistent with those of real lesions. We implement several methods on MatchNet [5] to compare including densifying supervision by randomly locations to synthesize (*Location inconsistent*), by Cycle-GAN to synthesize (*Appearance inconsistent*). To validate the visually realistic appearance of synthesized lesions, we also implement a supervision-densifying methods by simply placing calcification-like white dots across two views (*Unrealistic lesion*).
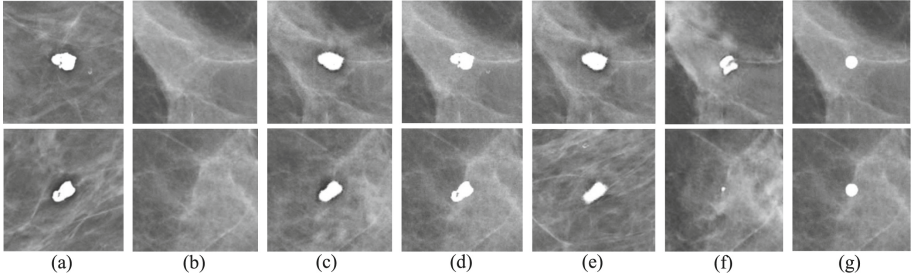

Fig. 3. Visualization of lesion generation using different methods. (a) and (b) are pairwise lesion and background GTs. (c) and (d) are respectively generated using our proposed approach and lesion-pasting, while (e)–(g) are respectively location inconsistent, appearance inconsistent and unrealistic lesion samples.

Visual examples of these four methods are shown in Fig. 3. Note that the locations of all examples except Fig. 3(e) are selected by our Gaussian model, which guarantees a correct spatial correlation between two views. Table 1 gives comparison results of the four methods, and our method achieves the best results for both b-Acc and AUC. By comparing b-ACC value of our method with the 1st and 2nd rows, we can see that 1) a correct spatial location can result in 1.971% improvement, and 2) appearance consistency can result in 2.948% improvement. By comparing the 2nd and 3rd rows, we surprisingly find that Cycle-GAN is even

worse than placing calcification-like dots by 2.094% of b-ACC, which implies that the inconsistency of appearance dramatically confused the DVT learning, resulting in unreasonable performance of cross-view matching.

Table 1. An ablation study of our proposed supervision-densifying method on INbreast. We report average results of balanced-Acc (%) and AUC.

Method	b-Acc (%)	AUC
Location inconsistent	79.631	0.888
Lesion inconsistent	78.654	0.888
Unrealistic lesion	80.748	0.892
Ours	**81.602**	**0.894**

Table 2. Quantitative comparison of different DVT-learning methods on both INbreast and CBIS-DDSM dataset. We evaluate on three general architectures and use ResNet-101 [6] as their unified feature network.

Method	MatchNet [5]		2-channel [20]		SCFM [13]	
	b-Acc(%)	AUC	b-Acc(%)	AUC	b-Acc(%)	AUC
Dataset: INbreast						
Sparse supervision	77.651	0.853	77.900	0.838	77.119	0.851
Lesion-pasting	80.200	0.890	77.840	0.848	80.103	0.881
Ours	**81.602**	**0.894**	**79.085**	**0.859**	**81.179**	**0.891**
Dataset: CBIS-DDSM						
Sparse supervision	72.904	0.796	70.664	0.776	71.686	0.788
Lesion-pasting	73.805	0.823	68.488	0.749	72.401	0.789
Ours	**76.092**	**0.839**	**73.529**	**0.812**	**76.294**	**0.824**

3.3   Comparison with Other DVT Learning Methods

Table 2 shows final results of different DVT learning methods, including training under sparse supervision, lesion-pasting and our proposed denser supervision. Lesion pasting directly crops areas of real lesions using masks and paste on pure backgrounds. Note that lesion-pasting requires a GT mask of each lesion while our method is trained in a weakly-supervised fashion and only need the location of each lesion. To eliminate the negative impact of less training sample by sparse supervision, we partially select true and false lesion patches in each view for lesion-pasting and ours to make the number of pairs for training equal. As shown in Table 2, we implement three state-of-the-art methods [5,13,20] for comparison, each of which is trained three times based on sparse supervision, lesion-pasting and our method respectively.

Two observations can be made from the results. First, sparse supervision achieves the worse performance, indicating the effectiveness of densified supervision for DVT learning. More specifically, our method greatly surpasses 3.951% (3.188%) of b-ACC and 0.041 (0.034) of AUC (1st and 2nd columns) to the state-of-the-art methods of mammogram matching adopting MatchNet [1] under sparse supervision on INbreast (CBIS-DDSM). Second, comparing with lesion pasting, our proposed approach improves even greater. Despite visually similar of both methods as shown in Fig. 3(c) and (d), without employing adversarial techniques, lesion pasting fails to consider lesion-surrounding tissue variation (e.g., shade) which possibly acts as an important clue of DVT. Moreover, comparing with lesion pasting utilizing pixel-wise masks, our supervision-densifying method only exploits weaker supervision of lesion locations, which shows more convenience in clinical usage and greater universality on various tasks.

We have calculated statistical significance of our proposed method compared to sparse supervision and lesion pasting. On INbreast, improvements by our method are mostly significant on MatchNet and SCDM ($p<0.05$ except for AUC vs. lesion pasting), while on CBIS-DDSM, our method is significant on all baselines ($p<0.05$ except for b-Acc vs. lesion pasting on 2-channel network).

4   Conclusions

In this paper, we propose a novel supervision-densifying approach towards a robust DVT learning. Those supervision-densified synthetic lesions remain both correct correlation of spatial locations by a Gaussian model and appearance consistency across two views by our proposed LT-GAN. Experiment results demonstrate that our densified supervision strongly contributes to DVT learning and performs superiority in cross-view μC matching. Future works will explore more advanced location correspondence measuring techniques and pairwise lesion synthesis architectures.

References

1. Balntas, V., Johns, E., Tang, L., Mikolajczyk, K.: Pn-net: Conjoined triple deep network for learning local image descriptors. arXiv preprint arXiv:1601.05030 (2016)
2. Beatty, J.: The radon transform and the mathematics of medical imaging (2012)
3. Engeland, S.V., Timp, S., Karssemeijer, N.: Finding corresponding regions of interest in mediolateral oblique and craniocaudal mammographic views. Med. Phys. **33**(9), 3203–3212 (2006)

4. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.: Improved training of wasserstein gans. arXiv preprint arXiv:1704.00028 (2017)
5. Han, X., Leung, T., Jia, Y., Sukthankar, R., Berg, A.C.: Matchnet: unifying feature and metric learning for patch-based matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3279–3286 (2015)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
7. Koch, G., Zemel, R., Salakhutdinov, R.: Siamese neural networks for one-shot image recognition. In: ICML Deep Learning Workshop, vol. 2. Lille (2015)
8. Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gorovoy, M., Rubin, D.L.: A curated mammography data set for use in computer-aided detection and diagnosis research. Sci. Data **4**(1), 1–9 (2017)
9. Ma, J., Li, X., Li, H., Wang, R., Menze, B., Zheng, W.S.: Cross-view relation networks for mammogram mass detection. In: 2020 25th International Conference on Pattern Recognition (ICPR), pp. 8632–8638. IEEE (2021)
10. Melekhov, I., Kannala, J., Rahtu, E.: Siamese network features for image matching. In: 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 378–383. IEEE (2016)
11. 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)
12. Perek, S., Hazan, A., Barkan, E., Akselrod-Ballin, A.: Siamese network for dual-view mammography mass matching. In: Stoyanov, D., et al. (eds.) RAMBO/BIA/TIA -2018. LNCS, vol. 11040, pp. 55–63. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00946-5_6
13. Quan, D., Fang, S., Liang, X., Wang, S., Jiao, L.: Cross-spectral image patch matching by learning features of the spatially connected patches in a shared space. In: Jawahar, C.V., Li, H., Mori, G., Schindler, K. (eds.) ACCV 2018. LNCS, vol. 11362, pp. 115–130. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20890-5_8
14. 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
15. Simo-Serra, E., Trulls, E., Ferraz, L., Kokkinos, I., Fua, P., Moreno-Noguer, F.: Discriminative learning of deep convolutional feature point descriptors. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 118–126 (2015)
16. Wilms, M., Krüger, J., Marx, M., Ehrhardt, J., Bischof, A., Handels, H.: Estimation of corresponding locations in ipsilateral mammograms: a comparison of different methods. In: Medical Imaging 2015: Computer-Aided Diagnosis, vol. 9414, p. 94142B. International Society for Optics and Photonics (2015)
17. Yan, Y., Conze, P.H., Lamard, M., Quellec, G., Cochener, B., Coatrieux, G.: Multitasking siamese networks for breast mass detection using dual-view mammogram matching. In: International Workshop on Machine Learning in Medical Imaging, pp. 312–321. Springer (2020)
18. Yang, H., Ciftci, U., Yin, L.: Facial expression recognition by de-expression residue learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2168–2177 (2018)

19. Yang, Z., et al.: MommiNet: mammographic multi-view mass identification networks. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12266, pp. 200–210. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59725-2_20
20. Zagoruyko, S., Komodakis, N.: Learning to compare image patches via convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4353–4361 (2015)
21. 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)