A PARAMETERIZED GENERATIVE ADVERSARIAL NETWORK USING CYCLIC PROJECTION FOR EXPLAINABLE MEDICAL IMAGE CLASSIFICATIONS

Xiangyu Xiong[1], Yue Sun[1], Xiaohong Liu[2], Chan-Tong Lam[1], Tong Tong[3], Hao Chen[4]
Qinquan Gao[3], Wei Ke[1], Tao Tan[1,]*

[1] Faculty of Applied Sciences, Macao Polytechnic University
[2] John Hopcroft Center (JHC) for Computer Science, Shanghai Jiao Tong University
[3] College of Physics and Information Engineering, Fuzhou University
[4] Department of Mathware, Jiangsu JITRI Sioux Technologies Co., Ltd

ABSTRACT

Although current data augmentation methods are successful to alleviate the data insufficiency, conventional augmentation are primarily intra-domain while advanced generative adversarial networks (GANs) generate images remaining uncertain, particularly in small-scale datasets. In this paper, we propose a parameterized GAN (ParaGAN) that effectively controls the changes of synthetic samples among domains and highlights the attention regions for downstream classification. Specifically, ParaGAN incorporates projection distance parameters in cyclic projection and projects the source images to the decision boundary to obtain the class-difference maps. Our experiments show that ParaGAN can consistently outperform the existing augmentation methods with explainable classification on two small-scale medical datasets.

Index Terms— Data augmentation, parameterized generative adversarial network, projection distance, explainable classification, small-scale datasets

1. INTRODUCTION

Deep neural networks have achieved success in computer vision fields [1, 2, 3, 4, 5, 6], where a large-scale dataset is crucial for effectively training. However, training neural networks on the small-scale datasets leads to the overfitting and poor generalization. Although regularization techniques [7, 8, 9] have been developed to prevent overfitting, augmenting training data is an effective option to address the data insufficiency.

Researchers have developed two main data augmentation methods in the past decade. Conventional augmentation method can explore prior knowledge [10], including random cropping, flipping, etc. However, its intra-domain augmented data contributes little to the description of decision boundaries for downstream classifications. Generative adversarial network (GAN) [11] and its variations have also emerged as augmentation means [12, 13]. However, the synthetic images are quality uncertain and variety insufficiency because of the mode collapsing [14]. Moreover, there is still uncertainty regarding the domain labels of the synthetic images even if soft labels are used [15].

Auxiliary information have been integrated with GANs to synthesize images of a given specific type. cGAN [16] generates images with a specific condition of class labels. AC-GAN [17] and VACGAN [18] introduce an auxiliary classifier to reconstruct the class labels. Later, this idea is expanded to cross-domain translation by reconstructing target domain labels, such as cCycleGAN [19, 20] and StarGAN [21]. However, the class label and domain label cannot make synthetic samples adjust the decision boundary for classifications.

To overcome these issues, we propose a parameterized GAN (ParaGAN) enabling controlling the degree variation of synthetic images. This is achieved by leveraging target-domain samples' distances to the optimal hyperplane as controllable parameters in forward path and considering source images' distances to the hyperplane in the backward reconstruction path. Experiments show that ParaGAN consistently outperforms the state-of-the-arts and provides a more transparent explanation than Grad-CAM [22].

Our main contributions are as follows: 1. A novel cyclic parameterized projection perpendicular to hyperplane controlling the variety of synthetic images. 2. An online augmentation manner with weighted synthetic loss enhancing binary classification. 3. A novel class-difference map enabling explaining the downstream classification.

2. MATERIALS AND METHODS

2.1. Dataset Acquisition and Evaluation Metrics

We collect a mixed breast ultrasound datasets which contains BUSI [23] and UDIAT [24] and a COVID-19 Dataset (COVID-CT) [25]. Accuracy (ACC) and the area under the receiver operating characteristic curve are used for evaluation.
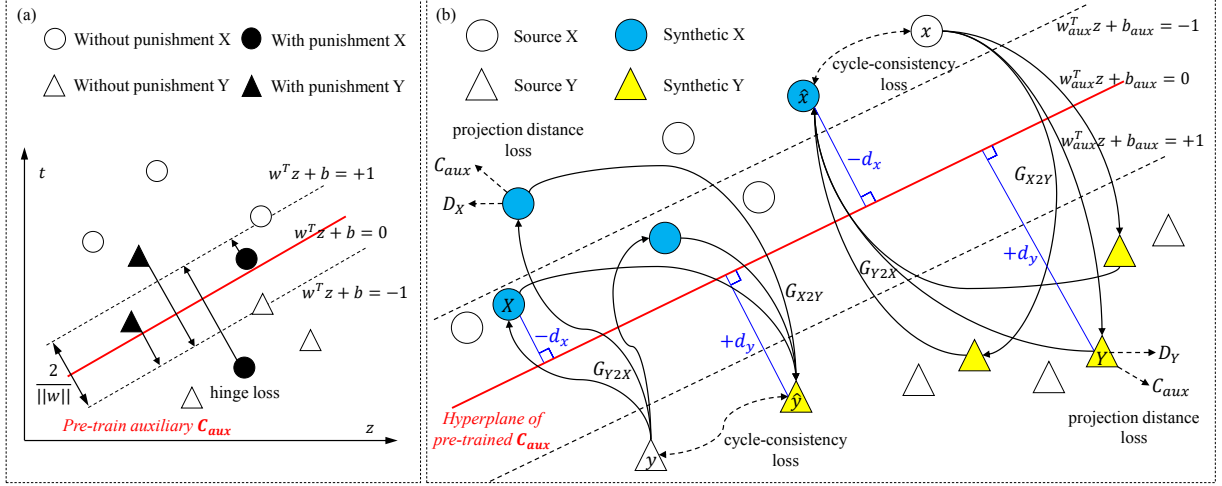
---

*Corresponding author

Fig. 1. Overview of ParaGAN enabling cyclic parameterized projection. (a) Pre-train an auxiliary classifier C_{aux} by hinge loss to provide a hyperplane. (b) The generators translate source images conditioned on the target images' projection distances in forward path, and vice versa for reconstructing sources. C_{aux} reconstructs the projection distances from synthetic images.

2.2. Parameterized Generative Adversarial Network

We pre-train a binary classifier by hinge loss to obtain an optimal hyperplane dividing the two domain samples, as shown in Figure 1(a). For a linear binary classifier $t = w^T z + b$, given a training set $\{z_i\}_{i=1}^N$, $z_i \in R^D$, $t_i \in \{-1, +1\}$, the hinge loss is defined as follows:

$$\mathcal{L}_{\text{hinge}}(t, \hat{t}) = \frac{1}{N} \sum_{i=1}^N max[0, 1 - t_i(w^T z_i + b))], \quad (1)$$

We develop parameterized generators G_{X2Y} and G_{Y2X} taking the projection distances to hyperplane as input parameters. To achieve this, we use pre-trained C_{aux} to measure projection distances from the real images to the optimal hyperplane ($w_{aux}^T z + b_{aux} = 0$). Given two random real images x and y, their vertical projection distances d_x and d_y are as:

$$d_x = |C_{aux}(x)|, x \in X; d_y = |C_{aux}(y)|, y \in Y. \quad (2)$$

Figure 1(b) illustrates the synthesis phase. Firstly, G_{X2Y} translates a source image x into $G_{X2Y}(x, +d_y)$ and D_Y distinguishes between this translated image and the real image y, and vice versa for G_{Y2X} and D_X. Secondly, an auxiliary classifier C_{aux} reconstructs the vertical projection distances d_y and d_x from the translated images $G_{X2Y}(x, +d_y)$ and $G_{Y2X}(y, -d_x)$, respectively. Thirdly, the source image x is reconstructed from the translated image $G_{X2Y}(x, +d_y)$ using its projection distance d_x, and vice versa for source y.

2.3. Objective Function

In the proposed ParaGAN, the adversarial loss for G_{X2Y} and D_Y is expressed as:

$$\begin{aligned} \mathcal{L}_{\text{GAN}}(G_{X2Y}, D_Y, X, Y) &= \mathbb{E}_{y \in Y}[log D_Y(y)] \\ &+ \mathbb{E}_{x \in X}[log(1 - D_Y(G_{X2Y}(x, +d_y)))], \end{aligned} \quad (3)$$

and vice versa for G_{Y2X} and D_X.

We propose a projection distance loss to force the synthetic images having same vertical distances as the target images. The projection distance loss can be formulated as:

$$\begin{aligned} &\mathcal{L}_{\text{proj}}(G_{X2Y}, G_{Y2X}, C_{aux}) \\ &= \mathbb{E}_{x \in X}[||(+d_y) - C_{aux}(G_{X2Y}(x, +d_y))||_2^2] \\ &+ \mathbb{E}_{y \in Y}[||(-d_x) - C_{aux}(G_{Y2X}(y, -d_x))||_2^2]. \end{aligned} \quad (4)$$

Cycle consistency loss is introduced to establish relationships between individual input x_i and a desired output y_i:

$$\begin{aligned} &\mathcal{L}_{\text{cyc}}(G_{X2Y}, G_{Y2X}) \\ &= \mathbb{E}_{x \in X}[||x - G_{Y2X}(G_{X2Y}(x, +d_y), -d_x)||_1] \\ &+ \mathbb{E}_{y \in Y}[||y - G_{X2Y}(G_{Y2X}(y, -d_x), +d_y)||_1]. \end{aligned} \quad (5)$$

Finally, the objective function for cross-domain synthesis represented by the following equation:

$$\begin{aligned} &\mathcal{L}_{\text{ParaGAN}}(G_{X2Y}, G_{Y2X}, C_{aux}, D_X, D_Y) \\ &= \mathcal{L}_{\text{GAN}}(G_{X2Y}, D_Y, X, Y) + \mathcal{L}_{\text{GAN}}(G_{Y2X}, D_X, Y, X) \\ &+ \lambda_{\text{proj}}\mathcal{L}_{\text{proj}}(G_{X2Y}, G_{Y2X}, C_{aux}) \\ &+ \lambda_{\text{cyc}}\mathcal{L}_{\text{cyc}}(G_{X2Y}, G_{Y2X}). \end{aligned} \quad (6)$$

where λ_{proj} and λ_{cyc} are weights that control the relative importance of projection distance loss and cycle consistency loss, respectively, compared to the adversarial loss.

2.4. Downstream Classifier Optimization

The loss of synthetic images in downstream classification is multiplied by a hyperparameter α, which controls the relative importance of synthetic data compared to true images. The combined loss of the downstream classifier can be defined as:

$$\mathcal{L}_{\text{C}}(c, \hat{c}) = \mathcal{L}_{\text{hinge}}(c, \hat{c}_{real}) + \alpha \mathcal{L}_{\text{hinge}}(c, \hat{c}_{syn}). \quad (7)$$
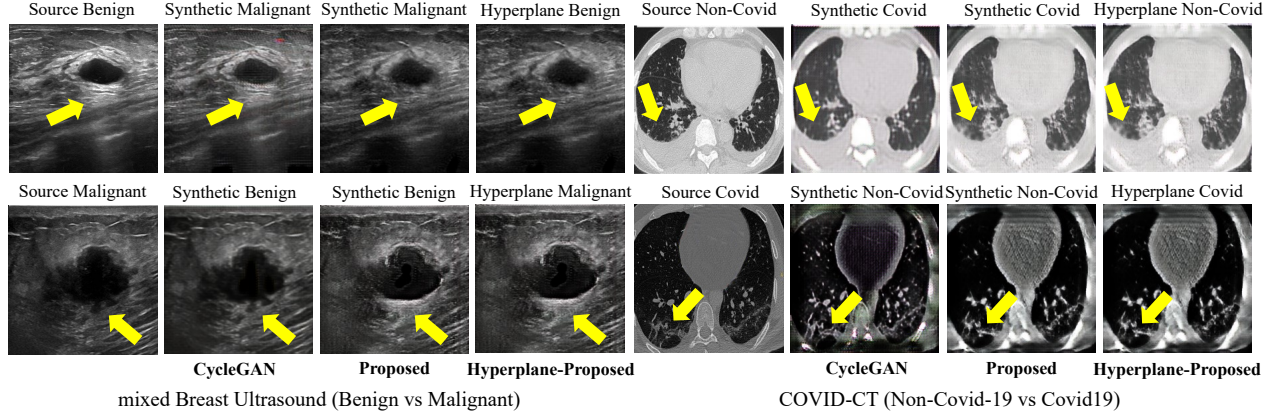
7311

Fig. 2. Qualitative results over mixed Breast Ultrasound and COVID-CT. The synthetic images by the ParaGAN are clearly closer to target domain than that by the CycleGAN [26]. The projected results on the hyperplane for source images are displayed.

Table 1. Comparison with the state of the arts over mixed Breast Ultrasound and COVID-CT. The proposed ParaGAN outperforms the state-of-the-arts using limited clinical data.

Methods	Loss	mixed Breast Ultrasound		COVID-CT	
		ACC	AUC	ACC	AUC
Original	CrossEntropy	0.849±0.000	0.928±0.013	0.767±0.049	0.836±0.043
Original	HingeLoss	0.887±0.025	0.933±0.019	0.757±0.079	0.840±0.071
Conventional Augmentation (CA)	CrossEntropy	0.877±0.016	0.933±0.003	0.765±0.015	0.845±0.027
Conventional Augmentation (CA)	HingeLoss	0.862±0.030	0.934±0.013	**0.796±0.010**	0.874±0.013
CA + ACGAN	HingeLoss	0.824±0.005	0.916±0.010	0.777±0.031	0.863±0.012
CA + VACGAN	HingeLoss	0.843±0.005	0.929±0.010	0.772±0.032	0.829±0.039
CA + CycleGAN	HingeLoss	0.849±0.028	0.915±0.003	0.785±0.017	0.866±0.017
CA + Proposed	HingeLoss	**0.895±0.048**	**0.947±0.017**	0.796±0.003	**0.883±0.027**

Table 2. Comparison with weighted synthetic losses (α) over mixed Breast Ultrasound and COVID-CT. Table shows optimal values of α are 0.2 and 1.0 for two datasets.

weighted synthetic loss α	mixed Breast Ultrasound		COVID-CT	
	ACC	AUC	ACC	AUC
0.2	**0.895±0.048**	**0.947±0.017**	0.793±0.021	0.874±0.021
0.4	0.871±0.011	0.938±0.020	0.777±0.017	0.880±0.005
0.6	0.855±0.022	0.936±0.014	0.782±0.012	0.861±0.002
0.8	0.862±0.038	0.938±0.008	**0.806±0.020**	0.858±0.021
1.0	0.846±0.005	0.940±0.016	0.796±0.003	**0.883±0.027**

Where c is the respective labels, $\hat{c}_{real}$ and $\hat{c}_{syn}$ are the downstream classifier's outputs for real data and synthetic data respectively, α is the loss weight of the synthetic data.

3. IMPLEMENTATION

3.1. Network Architecture

The generator's architecture is adopted from CycleGAN [26]. We add one channel for the first convolutional layer, because

the projection distance is required to spatially replicated match the size of the input image and concatenated with the input image. We use PatchGANs [27] as discriminator network to detect whether 70×70 overlapping image patches are real or synthetic. We adopt ConvNeXt [28] for the auxiliary classifier and the downstream classifier.

3.2. Training Settings

We train all networks with a learning rate of 0.0002 for the first 25 epochs and linearly decay the learning rate to 0 over the next 25 epochs. The hyperparameter λ_{cyc} is set to 10, and λ_{proj} is set to 0.1 considering the large scale of the distances. We adopt transfer learning for training all classifiers using ImageNet Dataset. The auxiliary classifier C_{aux} is not updated during the ParaGAN training procedure. We report ACC and AUC averaged over three runs.

4. EXPERIMENTS AND RESULTS

4.1. Experiments on Breast Ultrasound and COVID-CT

Table 1 quantitatively compares the proposed ParaGAN with several state-of-the-art GANs over datasets mixed Breast Ultrasound and COVID-CT. We can see that the proposed ParaGAN performs the best consistently especially when training samples are limited.

Figure 2 demonstrates that the proposed ParaGAN can generate more realistic and diverse images compared with the CycleGAN. We can see that the proposed ParaGAN makes the cross-domain changes including posterior regions of lesions in breast ultrasound image and the ground glass in the lungs CT image. Moreover, the source images and their projected images on hyperplane (Hyperplane-Proposed) are used to interpret the downstream classification in Section 4.4.
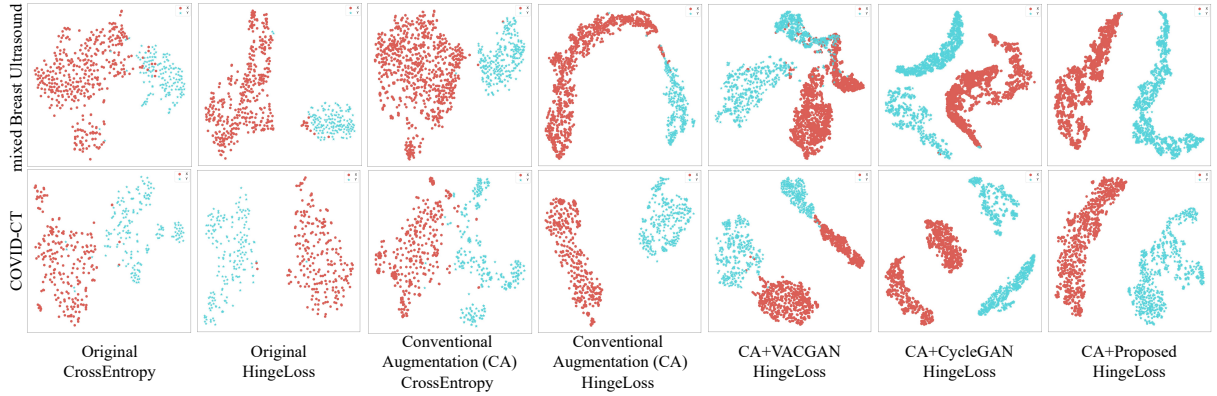
7312

Fig. 3. Training samples distributions of various methods for downstream ConvNext visualized by t-SNE [29]. The proposed ParaGAN tends to fill the latent space and to be along the margins of the hyperplane.
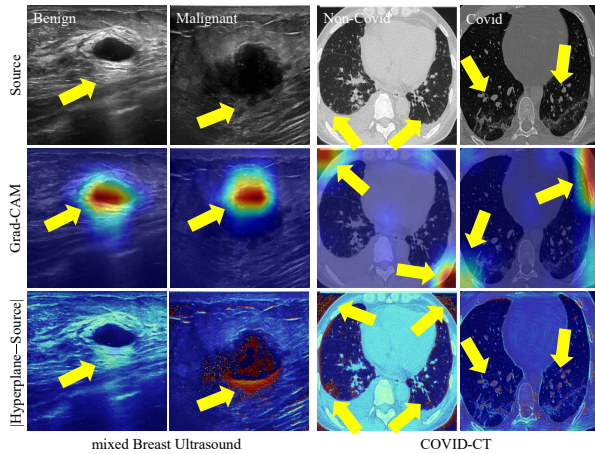

Fig. 4. The difference between the source images and their projections on a hyperplane can highlight class-specific region, which cannot be deduced from the Grad-CAM [22].

4.2. Ablation Study

The accuracy and auc can increase considerably by adjusting the synthetic weight (Table 2). With synthetic weights of 0.2 and 1.0, the auc of downstream classifier achieve the best performance. Using a synthetic weight allows for effective usage of GAN synthesized images, as synthetic images generally will not be as beneficial for a classifier as real images because of their faulty and somewhat unreliable nature.

4.3. Distribution of Training Samples

Figure 3 illustrates that how the cross-domain samples could help shape the decision boundary in the latent space from a downstream classifier. VACGAN converts noise-vector to images with condition of binary class labels, and CycleGAN translates source image to target image with only image space

constraint. These two methods lead to the label uncertainty of synthetic samples. The proposed ParaGAN translates source images to target images with both the image space constraint and the hyperplane distance constraint.

4.4. Interpretability of Classification

Figure 4 shows that ParaGAN provides the class-difference maps (CDMs) for explaining the binary classifier. The Grad-CAM focuses on the regions with large gradient changes in each class image, whereas the proposed class-difference maps (CDMs) focuses regions with large changes between each image and its projection on hyperplane. Specifically, we define the CDMs for images in domain X and domain Y using $|x - G_{X2Y}(x, 0)|$ and $|y - G_{Y2X}(y, 0)|$ respectively.

5. CONCLUSION AND FUTURE WORKS

We show that ParaGAN can generate samples to benefit and explain the downstream classification tasks, especially in small-scale datasets. We observe that controlling synthetic samples' variety have significantly more impact than blindly augmentation, and difference between the images and their projections on decision boundary can contribute explanation.

The current work has limitations that need to be studied in future. 1. Our work primarily focuses on the binary classification, and thus we will investigate levering hyperplane among multi-classes. 2. We will conduct a full-spectrum evaluation of the synthetic images in terms of clinical usefulness.

Acknowledgments

6. REFERENCES

[1] Rongsheng Wang et al., "Lightr-yolov5: A compact rotating detector for sars-cov-2 antigen-detection rapid diagnostic test results," *Displays*, vol. 78, pp. 102403, 2023.

[2] Rongsheng Wang et al., "Pctmf-net: heart sound classification with parallel cnns-transformer and second-order spectral analysis," *Vis. Comput.*, vol. 39, no. 8, pp. 3811–3822, 2023.

[3] Wenzhuo Zheng et al., "Multi-view 3d face reconstruction based on flame," *CoRR*, vol. abs/2308.07551, 2023.

[4] Shan Huang et al., "Transmrsr: transformer-based self-distilled generative prior for brain MRI super-resolution," *Vis. Comput.*, vol. 39, no. 8, pp. 3647–3659, 2023.

[5] Bingxin Wei et al., "Sofnet: Optical-flow based large-scale slice augmentation of brain mri," *Displays*, vol. 80, pp. 102536, 2023.

[6] Rongsheng Wang et al., "Ivygpt: Interactive chinese pathway language model in medical domain," *CoRR*, vol. abs/2307.10512, 2023.

[7] Nitish Srivastava et al., "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[8] Sergey Ioffe et al., "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. on Mach. Learn.*, 2015, vol. 37, pp. 448–456.

[9] Yuxin Wu et al., "Group normalization," in *Proc. Eur. Conf. on Comput. Vis.*, 2018, vol. 11217, pp. 3–19.

[10] Alex Krizhevsky et al., "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1106–1114.

[11] Ian J. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[12] Alec Radford et al., "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. Int. Conf. on Learn. Represent.*, 2016.

[13] M Frid-Adar et al., "Synthetic data augmentation using GAN for improved liver lesion classification," in *Proc. IEEE Int. Symp. Biomed. Imaging*, 2018, pp. 289–293.

[14] Divya Saxena et al., "Generative adversarial networks (gans): Challenges, solutions, and future directions," *ACM Comput. Surv.*, vol. 54, no. 3, pp. 63:1–63:42, 2022.

[15] Hongjiang Shi et al., "Data augmentation with improved generative adversarial networks," in *Prof. Int. Conf. on Pattern Recognit.*, 2018, pp. 73–78.

[16] Mehdi Mirza et al., "Conditional generative adversarial nets," *CoRR*, vol. abs/1411.1784, 2014.

[17] Augustus Odena et al., "Conditional image synthesis with auxiliary classifier gans," in *Proc. Int. Conf. on Mach. Learn.*, 2017, vol. 70, pp. 2642–2651.

[18] Shabab Bazrafkan et al., "Versatile auxiliary classifier with generative adversarial network (vac+gan), multi class scenarios," *CoRR*, vol. abs/1806.07751, 2018.

[19] Yongyi Lu et al., "Attribute-guided face generation using conditional cyclegan," in *Proc. Eur. Conf. on Comput. Vis.*, 2018, vol. 11216, pp. 293–308.

[20] Daichi Horita et al., "Food category transfer with conditional cyclegan and a large-scale food image dataset," in *Proc. CEA/MADiMa@IJCAI*, 2018, pp. 67–70.

[21] Yunjey Choi et al., "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proc. IEEE Int. Conf. on Comput. Vis. and Pattern Recognit.*, 2018, pp. 8789–8797.

[22] Ramprasaath R. Selvaraju et al., "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. on Comput. Vis.*, 2017, pp. 618–626.

[23] Walid Al-Dhabyani et al., "Dataset of breast ultrasound images," *Data in Brief*, vol. 28, pp. 104863, 2020.

[24] Moi Hoon Yap et al., "Automated breast ultrasound lesions detection using convolutional neural networks," *IEEE J. Biomed. Health. Inf.*, vol. 22, no. 4, pp. 1218–1226, 2018.

[25] Zhao et al., "Covid-ct-dataset: a ct scan dataset about covid-19," *arXiv preprint arXiv:2003.13865*, 2020.

[26] Junyan Zhu et al., "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. on Comput. Vis.*, 2017, pp. 2242–2251.

[27] Phillip Isola et al., "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Int. Conf. on Comput. Vis. and Pattern Recognit.*, 2017, pp. 5967–5976.

[28] Zhuang Liu et al., "A convnet for the 2020s," in *Proc. IEEE Int. Conf. on Comput. Vis. and Pattern Recognit.*, 2022, pp. 11966–11976.

[29] Van der Maaten et al., "Visualizing data using t-sne." *J Mach Learn Res*, vol. 9, no. 11, 2008.