

Please cite the paper as M.M. Rahman Siddiquee, J. Shah, T. Wu, C. Chong, T. Schwedt, and B. Li. HealthyGAN: Learning from Unannotated Medical Images to Detect Anomalies Associated with Human Disease. International Workshop on Simulation and Synthesis in Medical Imaging, 2022.

HealthyGAN: Learning from Unannotated Medical Images to Detect Anomalies Associated with Human Disease

Md Mahfuzur Rahman Siddiquee^{1,2}, Jay Shah^{1,2}, Teresa Wu^{1,2}, Catherine Chong^{2,3}, Todd Schwedt^{2,3}, and Baoxin Li^{1,2}

¹ Arizona State University, Tempe, AZ, USA
mrahmans@asu.edu

² ASU-Mayo Center for Innovative Imaging, Tempe, AZ, USA

³ Mayo Clinic, Phoenix, AZ, USA

Abstract. Automated anomaly detection from medical images, such as MRIs and X-rays, can significantly reduce human effort in disease diagnosis. Owing to the complexity of modeling anomalies and the high cost of manual annotation by domain experts (e.g., radiologists), a typical technique in the current medical imaging literature has focused on deriving diagnostic models from healthy subjects only, assuming the model will detect the images from patients as outliers. However, in many real-world scenarios, unannotated datasets with a mix of both healthy and diseased individuals are abundant. Therefore, this paper poses the research question of how to improve unsupervised anomaly detection by utilizing (1) an unannotated set of mixed images, in addition to (2) the set of healthy images as being used in the literature. To answer the question, we propose HealthyGAN, a novel one-directional image-to-image translation method, which learns to translate the images from the mixed dataset to only healthy images. Being one-directional, HealthyGAN relaxes the requirement of cycle-consistency of existing unpaired image-to-image translation methods, which is unattainable with mixed unannotated data. Once the translation is learned, we generate a difference map for any given image by subtracting its translated output. Regions of significant responses in the difference map correspond to potential anomalies (if any). Our HealthyGAN outperforms the conventional state-of-the-art methods by significant margins on two publicly available datasets: COVID-19 and NIH ChestX-ray14, and one institutional dataset collected from Mayo Clinic. The implementation is publicly available at <https://github.com/mahfuzmohammad/HealthyGAN>.

Keywords: Anomaly detection · COVID-19 detection · Thoracic disease detection · Migraine detection · Image-to-Image Translation.

1 Introduction

Supervised learning from a large annotated dataset is becoming easier [15, 10], due to deep neural networks. For problems like anomaly detection (e.g., rare

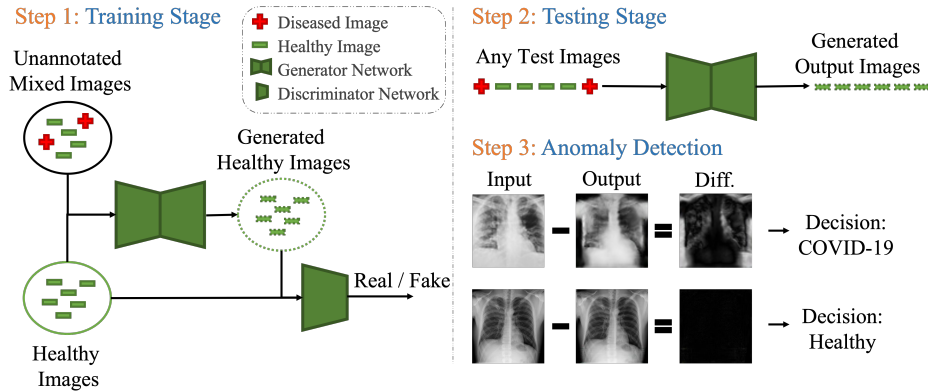


Fig. 1. Overview of the proposed anomaly detection method. At the training stage, our proposed HealthyGAN learns to generate healthy images utilizing an unannotated dataset mixed with both healthy and potential diseased/anomalous images, in addition to, a set of healthy images. At the testing stage, the absolute difference between the translated healthy image and the input image reveals the presence of an anomaly.

disease detection in medical images), however, it may often be challenging to obtain large enough datasets of annotated samples, making it impractical to rely on supervised learning for the task. Therefore, many recent attempts to develop diagnostic models learn only from the images of healthy (i.e. normal) subjects [7, 27, 28, 2, 35, 34, 1, 11, 25, 9, 26]. However, in practice, *unannotated* anomalous samples (mixed with normal samples) are usually available and what is missing is the elaborated annotation. In this research, we seek to answer the question: *How can we utilize an unannotated mixed dataset, in addition to the set of normal images, to improve the performance of anomaly detection?*

The answer to the question has been explored in the distant past for lesion detection in vascular CT images using SVM [40]. To the best of our knowledge, we have not seen any deep learning approach designed to address this question. Therefore, aiming to achieve a more generalized solution, in this paper we have developed a novel one-directional unpaired image-to-image translation network, termed HealthyGAN, based on Generative Adversarial Network [12, 13] (GAN). The proposed HealthyGAN learns to translate any images from a mixed dataset to normal images (a.k.a. healthy images) and detect abnormalities based on the differences between the input and the output images as illustrated in Fig. 1. We want to highlight that existing unpaired image-to-image translation methods [21, 30, 33, 39, 38, 8, 22, 36, 16, 20, 37] are not suitable for solving this problem since they are (1) bi-directional, requiring both abnormal-to-normal and normal-to-abnormal translation to ensure cycle-consistency [39]; or (2) require the images to be annotated as normal vs. abnormal *prior* to training. However, the translation of normal images back to abnormal images is not a feasible approach while using unannotated datasets. To address this challenge, HealthyGAN employs two important properties for improving anomaly detec-

tion: (1) unpaired image-to-image translation; and (2) one-directional image-to-image translation. To achieve these properties, we introduce a novel reconstruction loss that ensures effective cycle-consistency during the one-directional translation. Specifically, unlike traditional cycle-consistency loss [39], our reconstruction loss utilizes learned attention-masks to generate the reconstructed images for cycle-consistency. Since all the image manipulation for backward-cycle occurs using basic mathematical operations, there is no need for image annotation (see Sec. 2).

Through extensive experiments, we demonstrate that HealthyGAN outperforms existing state-of-the-art anomaly detection methods by significant margins on two public datasets: COVID-19 and NIH ChestX-ray14; and one single institute dataset for Migraine detection. This performance is attributed to HealthyGAN’s capability of utilizing unannotated diseased/anomalous images during training. In summary, we make the following contributions:

- We introduce a novel one-directional unpaired image-to-image translation method for anomaly detection that utilizes unannotated mixed datasets with images from healthy subjects and patients.
- We develop a novel reconstruction loss for ensuring cycle-consistency without requiring annotated inputs.
- With three challenging medical datasets, we perform extensive experiments comparing the proposed method, HealthyGAN, against the conventional state-of-the-art anomaly detection methods, and we report significant performance improvements and provide detailed analysis.

2 HealthyGAN: The Proposed Method

2.1 Network Architecture

The proposed HealthyGAN consists of a discriminator network and a generator network. The discriminator network follows PatchGAN [18, 19, 39] architecture and is similar to the ones used in [8, 24]. Our discriminator distinguishes whether the input image is a real or a fake (i.e. generated) healthy image.

The generator network takes any images without knowing their labels and translates them to only healthy images. For training, we use a mixed dataset (Set A) containing both diseased and healthy images, and another dataset (Set B) containing only healthy images. The generated images of these corresponding Sets are denoted as A' and B' , respectively. The generator does not generate the A' and B' images directly; rather, it generates intermediate healthy images B_{int} and masks M . The masks’ values are in the range $[0 - 1]$, where 0 means background pixel, and 1 means foreground pixel. Then we produce the final generated image B' following Eq. 1 and similarly, A' following Eq. 2.

$$B' = B_{int} \odot M + A \odot (\mathbf{1} - M) \quad (1)$$

$$A' = A \odot M + B_{int} \odot (\mathbf{1} - M) \quad (2)$$

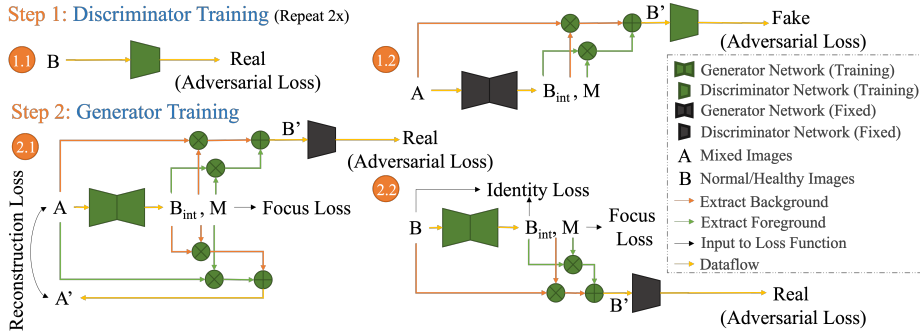


Fig. 2. A single training iteration of HealthyGAN, each of which consists of 2 sub-steps: (1) discriminator training and (2) generator training. The discriminator learns to distinguish real healthy images from the generated ones while the generator tries to fool the discriminator by generating realistic healthy images. Please see Sec. 2 for training details, Appx. A for hyper-parameters, and Appx. B for the network architectures.

If the image from Set A is diseased, we expect the mask M to activate the diseased region as foreground; otherwise, we expect M to be empty/zero. It is worth noting the similarity between Eq. 2 and the cycle-consistency concept introduced in [39]. Since the proposed method is controlling the image generation, partially, by the mask M , it neither requires a label nor an additional generator network to generate A' (Fig. 2). As the generator network translates the input images to a single direction, we call it a one-directional image-to-image translation method.

2.2 Training

Fig. 2 depicts the detailed training methodology of HealthyGAN. We train the generator and the discriminator network, alternately, like any GAN model. At each training step, we update the weights of the generator once for every two weight updates of the discriminator network and repeat until convergence.

The discriminator is trained to learn the healthy images in B as real and any images from the generator to be fake using an *adversarial loss* defined in Eq. 3.

$$\begin{aligned} \mathcal{L}_{adv}^D = & \mathbb{E}_{x \in A} [D_{real/fake}(G(x))] - \mathbb{E}_{x \in B} [D_{real/fake}(x)] \\ & + \lambda_{gp} \mathbb{E}_{\hat{x}} [(\|\nabla_{\hat{x}} D_{real/fake}(\hat{x})\|_2 - 1)^2] \end{aligned} \quad (3)$$

Here, $G(x)$ denotes the output of the generator and is obtained by Eq. 1. $D_{real/fake}(x)$ denotes the output of the discriminator network. Eq. 3 is the revised adversarial loss based on the Wasserstein GAN [3] and an added gradient penalty [14] with weight λ_{gp} which helps to stabilize the training.

The objective of the generator is to translate any input image to the corresponding healthy image. To be specific, if the input image is a healthy image, the generator is expected to behave like an autoencoder. If the input is a diseased image, the generator should remove anomalous parts and produce a healthy image in the output. The *adversarial loss* for the generator is defined in Eq. 4.

$$\mathcal{L}_{adv}^G = - \sum_{x \in \{A, B\}} \mathbb{E}_x [D_{real/fake}(G(x))] \quad (4)$$

For the known healthy image set, B , the generator should behave like an auto-encoder. Hence, we apply an *identity loss* (defined in Eq. 5) for these images.

$$\mathcal{L}_{id} = \mathbb{E}_{x \in B} [\|G_{int}(x) - x\|_1] \quad (5)$$

Here, G_{int} denotes the generated images before applying the masks (B_{int} in Fig. 2). Since we train HealthyGAN using unpaired images we add a *reconstruction loss* (Eq. 6) to ensure that the generated images are close to the input images.

$$\mathcal{L}_{rec} = \mathbb{E}_{x \in A, y \in A'} [\|x - y\|_1] \quad (6)$$

To control the size of the masks, we have adopted the *focus loss* of Eq. 7 from [23].

$$\mathcal{L}_f = \lambda_{fs} \left(\sum_{i=1}^n M_i/n \right)^2 + \lambda_{fz} \frac{1}{n} \sum_{i=1}^n \frac{1}{|M_i - 0.5| + \epsilon} \quad (7)$$

Here, n denotes the number of pixels in the mask M and M_i denotes a pixel in it. The first component controls the size of the mask and the second component forces the values to be close to 0/1. λ_{fs} and λ_{fz} are relative weights of these components, respectively.

Combining all losses, the final full objective function for the discriminator and generator can be described by Eq. 8 and Eq. 9, respectively.

$$\mathcal{L}_D = \mathcal{L}_{adv}^D \quad (8)$$

$$\mathcal{L}_G = \mathcal{L}_{adv}^G + \lambda_{rec} \mathcal{L}_{rec} + \lambda_{id} \mathcal{L}_{id} + \lambda_f \mathcal{L}_f \quad (9)$$

where λ_{rec} , λ_{id} , and λ_f determine the relative importance of the *reconstruction loss*, *identity loss*, and *focus loss*, respectively.

2.3 Detecting Anomalies

Fig. 1 provides an overview of the proposed anomaly detection method. Given an unannotated mixed dataset containing a mixture of both diseased and healthy images A and another dataset containing only healthy images B , we train the HealthyGAN as described in Sec. 2.2. Once trained, we first translate each of the test images into healthy images, and then we compute the absolute difference between the generated healthy images and the input images. We expect the resultant difference images to show the diseased regions if the input is a diseased image; otherwise, we expect the difference image to contain only pixels with a value of zero or very close to zero. Therefore, we detect the presence of the disease by checking the mean value of the difference images. Please note that the difference images indicate the presence of the disease/anomaly, and can also serve to localize image regions that associate with the disease/anomaly. However, the proposed HealthyGAN does not guarantee the detection of all the disease-specific features. Identifying only a subset of the disease-specific features is sufficient to serve the purpose of this study.

3 Experiments and Results

Competing Methods. We have compared the proposed HealthyGAN with 6 state-of-the-art anomaly detection methods currently in use. We have selected these methods as they are the most recent. Among them, ALAD [34], ALOCC [26], f-AnoGAN [28], and Ganomaly [1] are methodologically the closest to the proposed HealthyGAN. We have excluded other methodologically similar works such as EGBAD [35] and AnoGAN [27] from our competing methods’ list since ALAD and f-AnoGAN are improved versions of these methods, respectively. However, we have included PatchCore [25] and PaDiM [9], though methodologically different than the proposed HealthyGAN, as they are state of the art for novelty detection in natural image dataset like MVTec AD [4, 5].

Evaluation. We have compared the proposed HealthyGAN for anomaly detection with the conventional methods using the AUC score from the receiver operating characteristic (ROC) curve. In addition, we have reported precision, recall, specificity, and F1 scores. To get the prediction score for HealthyGAN, we first take the absolute difference between the input image and its translated image. Then we compute the mean value of the resultant difference image. We found the mean to be more robust than the maximum. For the conventional methods, we have used the anomaly score generation method proposed by their corresponding authors.

3.1 COVID-19 Detection

Dataset. We have utilized the COVIDx dataset from [31]. The original dataset contains a training set with 15,464 Chest X-rays (1,670 COVID-19 positives, 13,794 healthy) and a testing set with 200 Chest X-rays (100 positives and 100 healthy). For our experiments, we have randomly taken 10,031 healthy images for the known healthy training set. For the mixed unannotated training set, we have randomly taken 3,663 healthy and 1,570 COVID-19 positive images.

Results. The top section in Tab. 1 summarizes the COVID-19 detection results. As seen, HealthyGAN achieves COVID-19 detection AUC of 0.84 outperforming all the conventional methods by a large margin. It also achieves the best precision, specificity, and F1 scores of 0.76. In contrast, f-AnoGAN, the top performing among the competing methods, achieves an AUC score of only 0.64 which is 0.20 points lower than the proposed HealthyGAN. f-AnoGAN achieves precision, recall, specificity, and F1 scores of 0.55, 0.53, 0.56, and 0.54, respectively. Fig. 3 shows qualitative results of COVID-19 detection by HealthyGAN.

3.2 Chest X-ray 14 Diseases Detection

Dataset. We have utilized only the Posterior Anterior (PA) X-rays from the ChestX-ray14 dataset [32] for this experiment. The dataset contains X-rays with one or more of 14 thoracic diseases. For ease of evaluation, we selected the X-rays having only one disease. From the resultant X-rays, we used 10,000 healthy X-rays for the known healthy training set; 5,000 healthy X-rays, and 3,195 diseased

Table 1. Summary of the anomaly detection results. We have compared HealthyGAN with 6 state-of-the-art anomaly detection methods using 5 metrics on 3 medical imaging datasets. The best results are in **bold** and the second best results are underlined.

Datasets	Metrics	ALAD	ALOCC	f-AnoGAN	Ganomaly	Padim	PatchCore	HealthyGAN
COVID-19	AUC	0.58	0.63	<u>0.64</u>	0.58	0.56	0.52	0.84
	Prec.	0.49	<u>0.63</u>	0.55	0.59	0.56	0.52	0.76
	Rec.	0.89	0.63	0.53	0.60	0.56	0.53	<u>0.76</u>
	Spec.	0.09	<u>0.63</u>	0.56	0.59	0.56	0.51	0.76
	F1	<u>0.64</u>	0.63	0.54	0.60	0.56	0.52	0.76
X-ray 14 diseases	AUC	0.53	0.48	<u>0.55</u>	0.49	0.54	0.53	0.56
	Prec.	0.53	0.48	0.55	0.49	<u>0.54</u>	0.53	0.55
	Rec.	0.53	0.48	0.55	0.49	<u>0.54</u>	0.53	0.55
	Spec.	0.53	0.48	0.55	0.49	<u>0.54</u>	0.53	0.55
	F1	0.53	0.48	0.55	0.49	<u>0.54</u>	0.53	0.55
Migraine	AUC	0.60	0.40	0.50	<u>0.70</u>	0.35	0.60	0.75
	Prec.	0.60	0.40	0.50	<u>0.70</u>	0.36	0.60	0.78
	Rec.	0.60	0.40	0.50	<u>0.70</u>	0.40	0.60	0.70
	Spec.	0.60	0.40	0.50	<u>0.70</u>	0.30	0.60	0.80
	F1	0.60	0.40	0.50	<u>0.70</u>	0.38	0.60	0.74

X-rays for the mixed unannotated training set. For the validation set, we used 4,000 healthy and 4,000 diseased X-rays and for the testing set, we used 10,000 healthy and 10,000 diseased X-rays.

Results. The middle section of Tab. 1 summarizes the 14 diseases detection results. As seen, HealthyGAN achieves the best detection AUC score of 0.56 while f-AnoGAN performs the second best with an AUC score of 0.55. They both achieve precision, recall, specificity, and F1 scores of 0.55. Fig. 3 contains qualitative results of the 14 diseases detection by HealthyGAN.

3.3 Migraine Detection

Dataset. Our migraine dataset, collected by our collaborators at Mayo Clinic, contains 96 brain MRIs of migraine patients and 104 brain MRIs of healthy participants. We randomly selected 10 migraine patients and 10 healthy participants for each of the validation and test sets. The rest of the 76 migraine patients and 84 healthy participants were used as the mixed unannotated training set. For the known healthy set, we used 424 participants from the IXI public dataset [6].

Results. The last section of Tab. 1 summarizes the results for migraine detection. HealthyGAN outperforms all the conventional methods by achieving an AUC score of 0.75. It also achieves the best precision, recall, specificity, and F1 scores of 0.78, 0.70, 0.80, and 0.74, respectively. In contrast, the top performing conventional method, Ganomaly, achieves AUC, precision, recall, specificity, and F1 scores of only 0.70. Fig. 3 shows qualitative results of HealthyGAN.

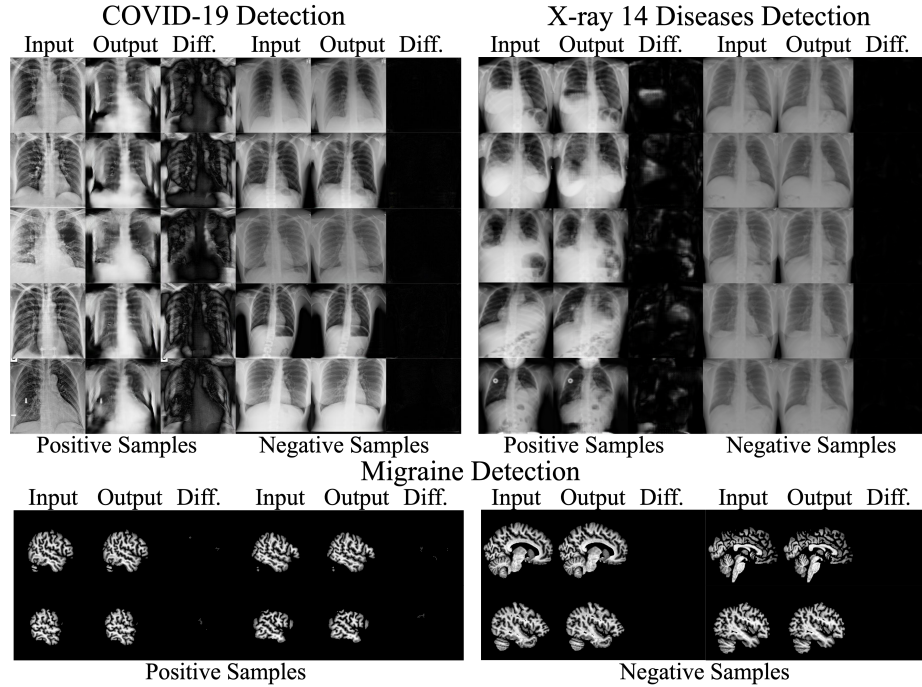


Fig. 3. Qualitative results of COVID-19, Chest X-ray 14 diseases detection, and Migraine detection by HealthyGAN. As seen, HealthyGAN has resulted in a high response in the difference maps for positive samples compared to the negative samples.

4 Conclusion

We have introduced a novel one-directional unpaired image-to-image translation method for anomaly detection from medical images, named HealthyGAN. We have devised a methodology to utilize an unannotated mixed dataset with both normal and anomalous images during the training of the proposed HealthyGAN. It has been possible due to the proposed novel reconstruction loss that ensures effective cycle-consistency without requiring input image annotations. Our extensive evaluation has demonstrated the proposed HealthyGAN’s superiority over the existing state-of-the-art anomaly detection methods. The superior performance is attributed to HealthyGAN’s capability of utilizing unannotated anomalous images during training.

Acknowledgments. This research has been supported by the United States Department of Defense W81XWH-15-1-0286 and W81XWH1910534, National Institutes of Health K23NS070891, National Institutes of Health - National Institute of Neurological Disorders and Stroke, Award Number 1R61NS113315-01, and Amgen Investigator Sponsored Study 20187183. We thank Arizona State University Research (ASURC) Computing for hosting our computing resources.

A Implementation Details

We have resized the input images to 256×256 for the experiments on COVID-19 and Migraine detection in Sec. 3.1 and Sec. 3.3, respectively. For the X-ray 14 diseases detection in Sec. 3.2, we have resized the images to 128×128 . We have set $\lambda_{gp} = 10$, $\lambda_{id} = 1$, $\lambda_{rec} = 1$, $\lambda_f = 0.1$, $\lambda_{fz} = 1$, and $\lambda_{fs} = 1$ for all the experiments. For COVID-19 and Migraine detection, we have used a batch-size of 16. For X-ray 14 diseases, we have used 32. We trained the models for 400,000 iterations. We have used Adam optimizer with a learning rate of $1e^{-4}$. The learning rate has been decayed for the last 100,000 iterations. Once trained, we have picked the best model using Fréchet inception distance (FID) [17, 29]. The network architecture details are provided in Appx. B.

B Network Architectures

B.1 Discriminator

Table 2. Discriminator network architecture. OC, KS, S, P, and NS stand for output channels, kernel size, stride, padding, and negative slope, respectively. The network architecture is adopted from [8, 24] with slight modification.

Type	Operations	Input Shape	Output Shape
Input layer	Conv2d (OC=64, KS=4, S=2, P=1), LeakyReLU (NS=0.01)	$(h, w, 3)$	$(\frac{h}{2}, \frac{w}{2}, 64)$
Hidden layers	Conv2d (OC=128, KS=4, S=2, P=1), LeakyReLU (NS=0.01)	$(\frac{h}{2}, \frac{w}{2}, 64)$	$(\frac{h}{4}, \frac{w}{4}, 128)$
	Conv2d (OC=256, KS=4, S=2, P=1), LeakyReLU (NS=0.01)	$(\frac{h}{4}, \frac{w}{4}, 128)$	$(\frac{h}{8}, \frac{w}{8}, 256)$
	Conv2d (OC=512, KS=4, S=2, P=1), LeakyReLU (NS=0.01)	$(\frac{h}{8}, \frac{w}{8}, 256)$	$(\frac{h}{16}, \frac{w}{16}, 512)$
	Conv2d (OC=1024, KS=4, S=2, P=1), LeakyReLU (NS=0.01)	$(\frac{h}{16}, \frac{w}{16}, 512)$	$(\frac{h}{32}, \frac{w}{32}, 1024)$
	Conv2d (OC=2048, KS=4, S=2, P=1), LeakyReLU (NS=0.01)	$(\frac{h}{32}, \frac{w}{32}, 1024)$	$(\frac{h}{64}, \frac{w}{64}, 2048)$
Output layer (D_{src})	Conv2d (OC=1, KS=3, S=1, P=1)	$(\frac{h}{64}, \frac{w}{64}, 2048)$	$(\frac{h}{64}, \frac{w}{64}, 1)$

B.2 Generator

Table 3. Generator network architecture. OC, KS, S, P, and IN stand for output channels, kernel size, stride, padding, and instance norm, respectively. The network architecture is adopted from [8, 24] with slight modification.

Type	Operations	Input Shape	Output Shape
Encoder	Conv2d (OC=64, KS=7, S=1, P=3), IN, ReLU	$(h, w, 3)$	$(h, w, 64)$
	Conv2d (OC=128, KS=4, S=2, P=1), IN, ReLU	$(h, w, 64)$	$(\frac{h}{2}, \frac{w}{2}, 128)$
	Conv2d (OC=256, KS=4, S=2, P=1), IN, ReLU	$(\frac{h}{2}, \frac{w}{2}, 128)$	$(\frac{h}{4}, \frac{w}{4}, 256)$
Bottleneck	Residual Block: Conv2d (OC=256, KS=3, S=1, P=1), IN, ReLU	$(\frac{h}{4}, \frac{w}{4}, 256)$	$(\frac{h}{4}, \frac{w}{4}, 256)$
	Residual Block: Conv2d (OC=256, KS=3, S=1, P=1), IN, ReLU	$(\frac{h}{4}, \frac{w}{4}, 256)$	$(\frac{h}{4}, \frac{w}{4}, 256)$
	Residual Block: Conv2d (OC=256, KS=3, S=1, P=1), IN, ReLU	$(\frac{h}{4}, \frac{w}{4}, 256)$	$(\frac{h}{4}, \frac{w}{4}, 256)$
	Residual Block: Conv2d (OC=256, KS=3, S=1, P=1), IN, ReLU	$(\frac{h}{4}, \frac{w}{4}, 256)$	$(\frac{h}{4}, \frac{w}{4}, 256)$
	Residual Block: Conv2d (OC=256, KS=3, S=1, P=1), IN, ReLU	$(\frac{h}{4}, \frac{w}{4}, 256)$	$(\frac{h}{4}, \frac{w}{4}, 256)$
	Residual Block: Conv2d (OC=256, KS=3, S=1, P=1), IN, ReLU	$(\frac{h}{4}, \frac{w}{4}, 256)$	$(\frac{h}{4}, \frac{w}{4}, 256)$
Decoder	ConvTranspose2d (OC=128, KS=4, S=2, P=1), IN, ReLU	$(\frac{h}{4}, \frac{w}{4}, 256)$	$(\frac{h}{4}, \frac{w}{4}, 128)$
	ConvTranspose2d (OC=64, KS=4, S=2, P=1), IN, ReLU	$(\frac{h}{2}, \frac{w}{2}, 128)$	$(h, w, 64)$
	ConvTranspose2d (OC=4, KS=7, S=1, P=3), Tanh	$(h, w, 64)$	$(h, w, 4)$

References

1. Akcay, S., *et al.*: Ganomaly: Semi-supervised anomaly detection via adversarial training. In: Asian conference on computer vision. pp. 622–637. Springer (2018)
2. Alex, V., *et al.*: Generative adversarial networks for brain lesion detection. In: Medical Imaging 2017: Image Processing. vol. 10133, p. 101330G. International Society for Optics and Photonics (2017)
3. Arjovsky, M., *et al.*: Wasserstein generative adversarial networks. In: International Conference on Machine Learning. pp. 214–223 (2017)
4. Bergmann, P., *et al.*: Mvtec ad—a comprehensive real-world dataset for unsupervised anomaly detection. In: Proceedings of the IEEE/CVF Conference on CVPR. pp. 9592–9600 (2019)
5. Bergmann, P., *et al.*: The mvtec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection. International Journal of Computer Vision **129**(4), 1038–1059 (2021)
6. Ixi dataset, <http://brain-development.org/ixi-dataset/>
7. Chen, X., *et al.*: Unsupervised detection of lesions in brain mri using constrained adversarial auto-encoders. arXiv preprint arXiv:1806.04972 (2018)
8. Choi, Y., *et al.*: Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In: The IEEE Conference on CVPR (June 2018)
9. Defard, T., *et al.*: Padim: A patch distribution modeling framework for anomaly detection and localization. In: International Conference on Pattern Recognition. pp. 475–489. Springer (2021)
10. Esteva, A., *et al.*: Dermatologist-level classification of skin cancer with deep neural networks. nature **542**(7639), 115–118 (2017)
11. Gherbi, E., *et al.*: An encoding adversarial network for anomaly detection. In: Asian Conference on Machine Learning. pp. 188–203. PMLR (2019)
12. Goodfellow, I., *et al.*: Generative adversarial nets. In: Advances in neural information processing systems. pp. 2672–2680 (2014)
13. Goodfellow, I., *et al.*: Generative adversarial networks. Communications of the ACM **63**(11), 139–144 (2020)
14. Gulrajani, I., *et al.*: Improved training of wasserstein gans. In: Advances in Neural Information Processing Systems. pp. 5767–5777 (2017)
15. He, K., *et al.*: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision. pp. 1026–1034 (2015)
16. He, Z., *et al.*: Attgan: Facial attribute editing by only changing what you want. IEEE Transactions on Image Processing **28**(11), 5464–5478 (2019)
17. Heusel, M., *et al.*: Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in neural information processing systems **30** (2017)
18. Isola, P., *et al.*: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on CVPR. pp. 1125–1134 (2017)
19. Li, C., *et al.*: Precomputed real-time texture synthesis with markovian generative adversarial networks. In: ECCV. pp. 702–716. Springer (2016)
20. Liu, M., *et al.*: Stgan: A unified selective transfer network for arbitrary image attribute editing. In: Proceedings of the IEEE Conference on CVPR. pp. 3673–3682 (2019)
21. Liu, M.Y., *et al.*: Unsupervised image-to-image translation networks. In: Advances in Neural Information Processing Systems. pp. 700–708 (2017)

22. Mejjati, Y.A., *et al.*: Unsupervised attention-guided image-to-image translation. In: Advances in Neural Information Processing Systems. pp. 3693–3703 (2018)
23. Nizan, O., *et al.*: Breaking the cycle-colleagues are all you need. In: Proceedings of the IEEE/CVF Conference on CVPR. pp. 7860–7869 (2020)
24. Rahman Siddiquee, M.M., *et al.*: Learning fixed points in generative adversarial networks: From image-to-image translation to disease detection and localization. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 191–200 (2019)
25. Roth, K., *et al.*: Towards total recall in industrial anomaly detection. arXiv preprint arXiv:2106.08265 (2021)
26. Sabokrou, M., *et al.*: Adversarially learned one-class classifier for novelty detection. In: Proceedings of the IEEE conference on CVPR. pp. 3379–3388 (2018)
27. Schlegl, T., *et al.*: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: International Conference on Information Processing in Medical Imaging. pp. 146–157. Springer (2017)
28. Schlegl, T., *et al.*: f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. Medical Image Analysis (2019)
29. Seitzer, M.: pytorch-fid: FID Score for PyTorch. <https://github.com/mseitzer/pytorch-fid> (August 2020), version 0.1.1
30. Shen, W., *et al.*: Learning residual images for face attribute manipulation. In: Proceedings of the IEEE conference on CVPR. pp. 4030–4038 (2017)
31. Wang, L., *et al.*: Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. Scientific Reports **10**(1), 19549 (Nov 2020)
32. Wang, X., *et al.*: Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE Conference on CVPR. pp. 2097–2106 (2017)
33. Yi, Z., *et al.*: Dualgan: Unsupervised dual learning for image-to-image translation. In: ICCV. pp. 2868–2876 (2017)
34. Zenati, H., *et al.*: Adversarially learned anomaly detection. In: 2018 IEEE International conference on data mining (ICDM). pp. 727–736. IEEE (2018)
35. Zenati, H., *et al.*: Efficient gan-based anomaly detection. arXiv preprint arXiv:1802.06222 (2018)
36. Zhang, G., *et al.*: Generative adversarial network with spatial attention for face attribute editing. In: Proceedings of the ECCV. pp. 417–432 (2018)
37. Zhao, Y., *et al.*: Unpaired image-to-image translation using adversarial consistency loss. arXiv preprint arXiv:2003.04858 (2020)
38. Zhu, J.Y., *et al.*: Toward multimodal image-to-image translation. In: Advances in Neural Information Processing Systems. pp. 465–476 (2017)
39. Zhu, J.Y., *et al.*: Unpaired image-to-image translation using cycle-consistent adversarial networks. arXiv preprint (2017)
40. Zuluaga, M.A., *et al.*: Learning from only positive and unlabeled data to detect lesions in vascular ct images. In: International conference on medical image computing and computer-assisted intervention. pp. 9–16. Springer (2011)