This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

CAPTURE-GAN: Conditional Attribute Preservation through Unveiling Realistic GAN for artifact removal in dual-energy CT imaging

Chunsu Park[1[0000−0002−9640−7619]], Seonho Kim[2[0009−0009−9565−0818]], DongEon Lee[2[0000−0002−0189−0231]], SiYeoul Lee[2[0009−0000−1817−6037]], Ashok Kambaluru[1[0009−0004−3031−3785]], Chankue Park[3[0000−0003−2937−114X]], and MinWoo Kim[1,4[0000−0001−7547−2596]⋆]

[1] School of Biomedical Convergence Engineering, College of Information and Biomedical Engineering, Pusan National University, Yangsan, Korea
[2] Department of Information Convergence Engineering, College of Information and Biomedical Convergence Engineering, Pusan National University, Yangsan, Korea
[3] Department of Radiology, Research Institute for Convergence of Biomedical Science and Technology, Pusan National University Yangsan Hospital, Yangsan, Korea
[4] Center for Artificial Intelligence Research, Pusan National University, Busan, Korea
mkim180@pusan.ac.kr

Abstract. Dual-energy CT (DECT) is gaining attention as an effective medical imaging modality for detecting bone marrow edema. However, imaging is complicated by the lower contrast offered by DECT compared to MRI and the inherent presence of artifacts in the image formation process, necessitating expertise in DECT. Despite advancements in AI-based solutions for image enhancement, achieving an artifact-free outcome in DECT remains difficult due to the impracticality of obtaining paired ground-truth and artifact-containing images for supervised learning. Recently, unsupervised techniques demonstrate high performance in image translation tasks. However, these methods face challenges in DECT due to the similarity between artifact and pathological patterns and could have a detrimental impact on image interpretation. In this study, we developed CAPTURE-GAN, which leverages a pre-trained classifier to preserve edema characteristics while removing DECT artifacts. Additionally, we introduced a mask indicating local regions pertaining to artifacts in order to prevent the output of the model from being over-smoothed or losing the bones' structural outline. Our approach fully utilizes automatically generated masks within the overall framework to only selectively modify the necessary local regions more cleanly and precisely than existing networks while preserving intricate bone patterns. Particularly, the performance of the classifier on artifact-removed images has been shown to surpass corresponding images before artifact removal. Code and models are available at https://github.com/pnu-amilab/CAPTURE-GAN.

Keywords: Artifact · Unsupervised learning · Bone marrow edema

⋆ Corresponding author: mkim180@pusan.ac.kr

1 Introduction

One of the pivotal applications of dual-energy CT (DECT) is the detection of bone marrow edema (BME), which is a precursor to fractures. Despite its lower contrast relative to MRI, DECT enables the visualization of fluid within bone through material decomposition techniques [5, 18]. However, artifacts that are an inevitable part of the image formation process can obscure or simulate the pathological patterns of fluids, leading to incorrect interpretations if not accurately identified. Consequently, diagnosing BME or precisely localizing lesions requires extensive expertise in DECT.

Recent advances in AI-based solutions have shown promise in assisting radiologists with their diagnostic decisions by enhancing the quality of images [1, 19]. These solutions predominantly use deep neural network frameworks, operating in a supervised manner with extensive datasets. However, achieving an artifact-free outcome in DECT is challenging, as it is impractical to obtain a ground-truth image that is paired with its artifact-containing counterpart for supervisory purposes.

A potential strategy involves the use of unsupervised techniques for artifact removal and subsequent inpainting of the affected areas. For instance, the CycleGAN architecture [20] enables domain translation using unpaired image datasets. Similarly, AttGAN [8] incorporates an embedded label for a specific attribute as input, facilitating attribute modification. While these networks have shown impressive performance in processing natural camera images, their efficacy in DECT imaging tasks is not assured. This uncertainty arises because the artifacts and the pathological patterns are markedly similar, and designating artifacts as a specific attribute for targeted removal could inadvertently alter the pattern of edema due to their resemblance. Consequently, efforts to enhance specificity may inadvertently reduce sensitivity.

In this study, we introduce the Conditional Attribute Preservation through Unveiling Realistic GAN (CAPTURE-GAN) framework, designed to minimize artifacts while preserving the pathology of BME and the anatomical integrity of bone. The foundational concept lies in integrating a generative model, inspired by CycleGAN, with conditional constraints applied through masking and classification models. Specifically, CycleGAN is employed to generate plausible artifact-free images from corrupted ones, utilizing a cycle consistency loss as its constraint mechanism. Nonetheless, in the demanding context of DECT imaging, where artifacts can mimic pathological patterns to the unaided eye, CycleGAN alone struggles to distinguish fine details, often leading to the obliteration of both the intricate internal bone patterns and the bone's structural outline. Similar to the research conducted by [17, 10], we have developed an automated masking technique that introduces bone priors into CycleGAN, enhancing its focus on preserving bone structure while selectively removing artifacts. Furthermore, integrating a disease classification network imposes additional constraints on CycleGAN, ensuring that the generated images do not obscure essential patterns critical for disease diagnosis.

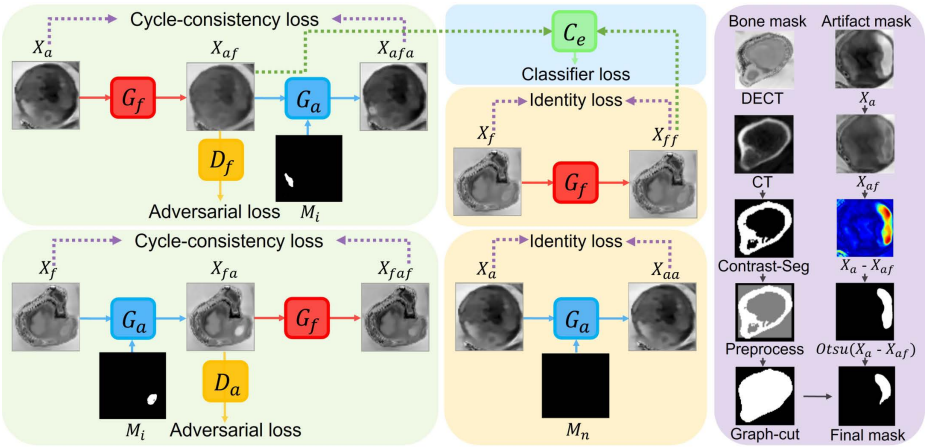


Fig. 1. CAPTURE-GAN Overview. This architecture, based on CycleGAN, includes an artifact-removal generator (G_f) and an artifact-corrupting generator (G_a), along with their associated adversarial discriminators D_f and D_a. The forward cycle translates images from the artifact-corrupted domain to the artifact-free domain and back again, while the backward cycle performs the reverse. The mask input to G_a assists the generator in synthesizing realistically corrupted images. When G_f and G_a receive artifact-free and artifact-corrupted images, respectively, they function as identity operators. The artifact-free images produced by G_f are assessed by the BME classifier C_e to determine whether G_f retains disease-distinguishing patterns.

2 Method

Figure 1 presents an overview of our model, CAPTURE-GAN, designed to selectively eliminate artifacts from DECT images. The architecture comprises multiple neural networks, including one classifier, two generators, two discriminators, and one non-neural-network-based mask creator. These components are strategically interconnected, enabling one generator (G_f) to access a diverse set of realistically augmented, artifact-corrupted bone images. This configuration directs the model to proficiently distinguish between artifacts and pathological BME patterns, facilitating effective isolation.

2.1 CAPTURE-GAN

Our CAPTURE-GAN model is derived from a CycleGAN framework. As depicted in Figure 1, in the forward cycle, an artifact-corrupted image X_a is fed into the artifact-removal generator G_f, which produces an artifact-free image $\hat{X}_{af}$. This clean image is then input into the artifact-corrupting generator G_a along with a mask M_i to generate a corrupted image $\hat{X}_{afa}$. The mask assists the network in reconstructing an image closely resembling X_a. The cycle-consistency loss, combined with an adversarial loss, compels the discriminator D_f to ensure that G_f preserves the bone outline and pathological patterns.

In the backward cycle, each artifact-free image X_f is introduced into G_a to produce a corrupted image $\hat{X}_{fa}$, again using the mask M_i. The mask's role here is to allow G_a to introduce varied plausible artifact patterns within the bone (as shown in supplementary). Processing this synthetic image through G_f results in a clean image $\hat{X}_{faf}$, enhancing the network's robustness through augmentation. The cycle-consistency and adversarial losses obligate G_a to create more realistic corrupted images.

Beyond these cycles, additional processes are also in place. When G_f and G_a receive X_f and X_a, respectively, they are expected to function as identity operators. Minimizing the difference between input and output helps preserve the identity. Here, G_a utilizes a zero-value mask M_n, to ensure that additional artifact generation is not present when creating $\hat{X}_a$ from X_a.

Our architecture expands on the work by [12], utilizing a ResUNet-based structure for both G_a and G_f. This architecture is characterized by several key features: 1) it is segmented into encoder and decoder sections; 2) it incorporates squeeze-and-excitation blocks [9] to highlight critical layer features; 3) it utilizes Atrous Spatial Pyramidal Pooling (ASPP) [3] to address both local and global features concurrently; and 4) it ensures the preservation of vital feature information through the inclusion of residual blocks [7]. The designs of D_a and D_f are based on the PatchGAN architecture [11], similar to [10].

2.2 Pre-trained classifier and mask creator

The CAPTURE-GAN model incorporates the binary classifier C_e, which differentiates between edema and normal images. The structure of classifier utilizes the ResNet18 architecture from [7], and it was pre-trained on bone images from 56 patient cases prior to the comprehensive training of the model. The feedback from C_e strengthens the capability of the artifact-removal generator G_f to preserve pathological details. The training of the classifier utilized the Adam optimizer [13] with a learning rate of 0.0002 and a mini-batch size of 16. To address the challenges of data scarcity and class imbalance, conventional augmentation techniques were applied. Validation on an image dataset independent from the training set yielded an accuracy of 88.7%.

Each binary mask M_i, utilized by the artifact-corrupting generator G_a, is automatically generated through the combination of two intermediate masks during model training. Figure 1 illustrates the process, where the first intermediate mask, the artifact mask, delineates the shapes of artifacts. This mask is derived during the forward cycle from the difference between the input and the output of the artifact-removal generator G_f, calculated as $U(\max(X_a - \hat{X}_{af}, 0), \theta)$, where $U(\cdot)$ represents the unit (binary) step function and θ denotes the adaptively determined threshold value by the Otsu method [16]. In the backward cycle, an artifact mask is chosen from a pool of all artifact masks generated during the forward cycle, enabling G_a to produce both realistic and varied artifact images $\hat{X}_{fa}$. The second intermediate mask, the bone mask, outlines the shapes of the femur bones and is extracted from X_a and X_f in the forward and backward

cycles, respectively. It is generated through initial contrast-based segmentation followed by graph-cut-based segmentation [2]. The purpose of this mask is to ensure that any artifacts are confined within the bone region, especially during the backward cycle. Consequently, the final mask M_i is created by intersecting the artifact and bone masks, ensuring precise artifact simulation within the bone structure.

2.3 Loss function

The loss function employed to update our model comprises multiple components. The adversarial loss for training the artifact-removal generator G_f is defined as follows:

$$L_{adv} = E_{X\sim P(X_f)}[\log D_f(X)] + E_{X\sim P(X_a)}[\log(1 - D_f(G_f(X)))]. \quad (1)$$

Meanwhile, the adversarial loss for updating the artifact-corrupting generator G_a is expressed as follows:

$$L_{adv} = E_{X\sim P(X_a)}[\log D_a(X)] + E_{X\sim P(X_f)}[\log(1 - D_a(G_a(X, M_i)))]. \quad (2)$$

The cycle-consistency loss is formulated for the dual generators as follows:

$$\begin{aligned} L_{cyc} = E_{X\sim P(X_f)}[||X - G_f(G_a(X, M_i))||_1] + \\ E_{X\sim P(X_a)}[||X - G_a(G_f(X), M_i)||_1]. \end{aligned} \quad (3)$$

When the generators G_f and G_a receive images X_f and X_a respectively, they should act as identity operators. To enforce this, the mask M_n used for G_a is configured as a zero matrix to signal no modification. The corresponding identity loss is given as follows:

$$L_{ident} = E_{X\sim P(X_f)}[||X - G_f(X)||_1] + E_{X\sim P(X_a)}[||X - G_a(X, M_n)||_1]. \quad (4)$$

The disease classifier C_e analyzes images produced by the artifact-removal generator G_f. Reducing the loss of C_e encourages G_f to preserve disease patterns while eliminating artifacts. Given C_e's softmax output layer, the classification loss is defined as follows:

$$L_{cls} = E_{X\sim P(X_a)}[-\mathbf{y}^T \log(\hat{\mathbf{y}})] + E_{X\sim P(X_f)}[-\mathbf{y}^T \log(\hat{\mathbf{y}})], \quad (5)$$

where $\hat{\mathbf{y}} = C_e(G_f(X)) \in \mathbb{R}^2$ denotes the prediction (probability) vector and $\mathbf{y} \in \mathbb{R}^2$ denotes the one-hot encoded label vector. The overall loss for CAPTURE-GAN is the weighted sum of these four losses is defined as follows:

$$L_{CAPTURE} = L_{adv} + \lambda_{cyc} L_{cyc} + \lambda_{ident} L_{ident} + \lambda_{cls} L_{cls}, \quad (6)$$

where $\lambda_{cyc} = 30$, $\lambda_{ident} = 15$ and $\lambda_{cls} = 1$ are weights determined empirically. Specifically, we evaluated 100 randomly selected images from the training set, choosing final weights that produced the most realistic images while maintaining bone structure and edema patterns, as shown in Supplementary Table 1.

2.4 Model training

We initialized the parameters of all networks using the Kaiming Normal method [6]. The Adam optimizer, with a learning rate of 0.0002, was employed for network optimization. The first and second momentum values were configured to 0.5 and 0.999, respectively. All networks underwent training for 500 epochs with a mini-batch size of 4, using a A100 GPU. For data augmentation purposes, a randomly cropped image of size 224×224 was used as training data at each iteration.

3 Experimental Results

3.1 Dataset

We collected DECT and corresponding MRI images from 70 subjects, adhering to criteria that included: the availability of MRI images for precise annotation, scans performed within a month of each other, and clear artifact identification. DECT scans were executed using 80 and 140 kVp settings (Revolution CT; GE Healthcare), comprising conventional and water-HAP axial reconstructed images. Regions of interest were marked, extracted, and resized to (256×256). For training, we used 1,837 artifact-corrupted slices and 1,064 artifact-free slices from 56 subjects. For testing, we used 504 artifact-corrupted slices and 227 artifact-free slices from the remaining 14 subjects.

3.2 Evaluation

We evaluated the performance of our proposed CAPTURE-GAN against a range of unsupervised attribute-editing image reconstruction methods, including FaderNet [14], AttGAN [8], StarGAN [4], CycleGAN [20], and STGAN [15]. To highlight the versatility of our model beyond the cycle consistency-based framework, we conducted experiments by replacing CycleGAN with AttGAN.

The ability of each model to preserve edema information was assessed. Artifact-corrupted images were processed by each model and then evaluated using a pre-trained disease classifier. This allowed for the comparison of diagnostic scores across different models. Additionally, artifact-free images were inputted into each model. The outputs were anticipated to be identical to the input images, thereby acting as ground truths. Consequently, we calculated peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and mean absolute error (MAE).

To gauge the artifact-removal efficacy, we employed pre-trained classifier to discern whether input images were artifact-free or artifact-corrupted. Following the methodology used with the disease classifier, we inputted the filtered images into the artifact-detecting classifier and examined the classification accuracies.

In our final analysis, we conducted an ablation study by incrementally adding components to the backbone. This study was aimed at uncovering the contributions of each component toward the overall performance.

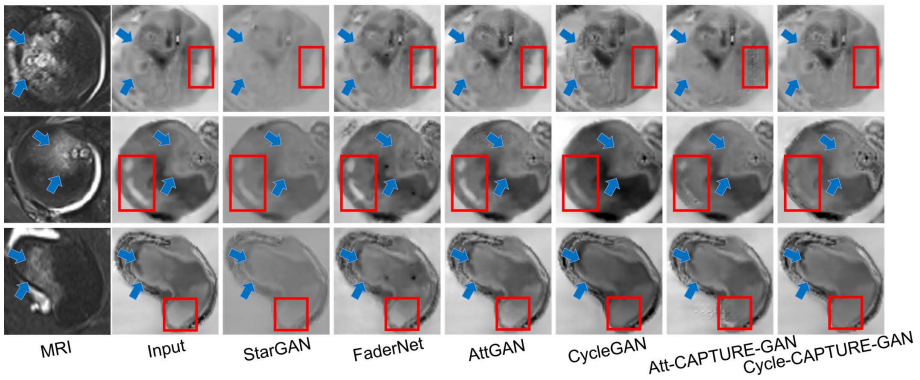


Fig. 2. Filtered images were produced from an artifact-corrupted input image by each respective model. The first and second columns display MRI and corresponding DECT input images, respectively. The subsequent columns show the output filtered images. Red boxes highlight artifacts, while blue arrows indicate pathological patterns.

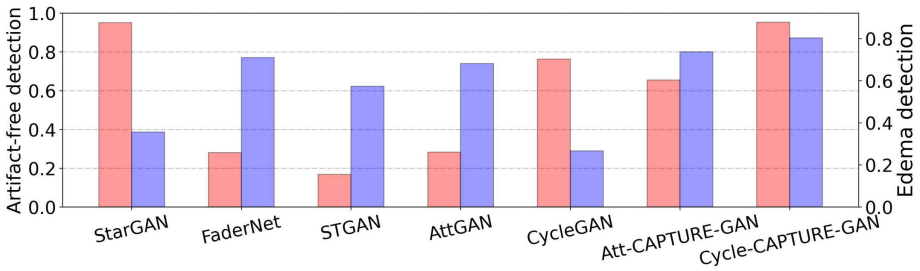
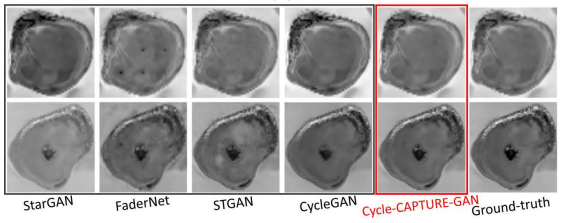


Fig. 3. Images corrupted by artifacts were processed by each model, and the resulting filtered images were then fed into both the artifact-detecting and disease-detecting classifiers. Red bars indicate the accuracies achieved in artifact-detection classification, while blue bars display the F1 scores from disease-detection classification outcomes.

3.3 Qualitative and Quantitative Comparison Results

Fig.2 displays the artifact filtering outcomes for each model when processing artifact-corrupted images. Compared to the input images, our model, CAPTURE-GAN, showcased superior efficacy in eliminating a variety of artifact patterns while preserving edema characteristics, outperforming other models. Concurrently, CAPTURE-GAN preserved the gray tissue structure and texture, crucial for disease identification, with minimal loss. These findings are corroborated by quantitative evaluation results. Fig.3 presents the outcomes of two classifiers. Our model demonstrated an exceptional balance between artifact removal and the preservation of pathological patterns. Moreover, Fig.4 (a) shows the output images when artifact-free images were used as inputs. This highlights the fact that our model preserved the intricate bone structure without any distor-

(a)



(b)

Method	PSNR ↑	SSIM ↑	MAE ↓
StarGAN	30.22	0.898	0.028
FaderNet	28.60	0.882	0.027
STGAN	34.70	0.961	0.012
CycleGAN	29.36	0.956	0.030
Cycle-CAPTURE-GAN	39.52	0.986	0.011

Fig. 4. (a) Images generated by each model. The output images were produced by each model from an artifact-free input image. The images in the left five columns are the output images, and the images in the rightmost column are the input images (ground-truths). (b) Quantitative comparisons between models. PSNR, SSIM and MAE between ground-truth and artifact-free images processed by each model were calculated.

Table 1. Ablation study results. The scores were obtained when artifact-corrupted images were processed by each model, and the resulting images were subsequently input into either the artifact-detecting classifier or the edema-detecting classifier.

Method	ACC (Artifact) (%)	ACC (Edema) (%)	AUC (Edema) (%)	F1 Score (Edema) (%)
None	-	88.10	93.74	75.00
AttGAN	25.99	82.24	92.52	68.10
Att-CAPTURE-GAN	60.32	90.08	91.63	73.68
CycleGAN	70.24	80.36	64.65	26.67
Cycle-CAPTURE-GAN (w/o mask)	81.35	88.49	95.12	77.52
Cycle-CAPTURE-GAN (w/ mask)	87.70	90.67	95.16	80.33

tions. Fig.4 (b), which summarizes the quantitative scores, further emphasizes the superiority of our model.

3.4 Ablation Study Results

Table 1 elucidates the impact of each function on artifact removal and the preservation of pathological patterns. As the baseline GAN structure, CycleGAN demonstrates more effective artifact removal than AttGAN but at the expense of a loss of pathological information. Implementing constraint techniques across any backbone significantly boosts performance. Integrating a disease classifier within CycleGAN, along with the addition of a masking technique, systematically diminishes the loss of edema patterns and augments the artifact removal capability. It is important to note that filtering significantly improved the performance of the edema classifier, resulting in notably higher diagnostic scores when evaluating filtered images compared to unfiltered images.

4 Conclusion

This study introduces an innovative framework designed to enhance Dual-Energy Computed Tomography (DECT) imaging by preserving critical edema patterns

while eliminating artifacts. This approach leverages a classifier and masks within a generative adversarial network, presenting a straightforward yet significant method for retaining essential edema information , a crucial aspect that must not be compromised. Impressively, our model surpasses existing networks in retaining edema features while effectively clearing artifacts. Our future research aims to examine the impact of artifact removal functions on medical practitioners' ability to identify local pathological features accurately. We are particularly interested in determining whether such functions can enhance the diagnostic accuracy of medical doctors with limited experience in interpreting DECT.

Acknowledgments. This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A2C2094778). This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) under the Artificial Intelligence Convergence Innovation Human Resources Development (IITP-2024-RS-2023-00254177) grant funded by the Korea government(MSIT). This work was supported by Institute of Information & communications Technology Planning & Evaluation(IITP) under the Leading Generative AI Human Resources Development(IITP-2024-RS-2024-00360227) grant funded by the Korea government(MSIT).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., Gatidis, S., Yang, B.: Medgan: Medical image translation using gans. Computerized medical imaging and graphics **79**, 101684 (2020)
2. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient nd image segmentation. International journal of computer vision **70**(2), 109–131 (2006)
3. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence **40**(4), 834–848 (2017)
4. Choi, Y., Choi, M., Kim, M., Ha, J.W., Kim, S., Choo, J.: Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8789–8797 (2018)
5. Foti, G., Catania, M., Caia, S., Romano, L., Beltramello, A., Zorzi, C., Carbognin, G.: Identification of bone marrow edema of the ankle: diagnostic accuracy of dual-energy ct in comparison with mri. La radiologia medica **124**, 1028–1036 (2019)
6. He, K., Zhang, X., Ren, S., Sun, J.: 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)
7. 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)

8. He, Z., Zuo, W., Kan, M., Shan, S., Chen, X.: Attgan: Facial attribute editing by only changing what you want. IEEE transactions on image processing **28**(11), 5464–5478 (2019)
9. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)
10. Hu, X., Jiang, Y., Fu, C.W., Heng, P.A.: Mask-shadowgan: Learning to remove shadows from unpaired data. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 2472–2481 (2019)
11. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017)
12. Jha, D., Smedsrud, P.H., Riegler, M.A., Johansen, D., De Lange, T., Halvorsen, P., Johansen, H.D.: Resunet++: An advanced architecture for medical image segmentation. In: 2019 IEEE international symposium on multimedia (ISM). pp. 225–2255. IEEE (2019)
13. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
14. Lample, G., Zeghidour, N., Usunier, N., Bordes, A., Denoyer, L., Ranzato, M.: Fader networks: Manipulating images by sliding attributes. Advances in neural information processing systems **30** (2017)
15. Liu, M., Ding, Y., Xia, M., Liu, X., Ding, E., Zuo, W., Wen, S.: Stgan: A unified selective transfer network for arbitrary image attribute editing. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3673–3682 (2019)
16. Otsu, N., et al.: A threshold selection method from gray-level histograms. Automatica **11**(285-296), 23–27 (1975)
17. Phan, V.M.H., Liao, Z., Verjans, J.W., To, M.S.: Structure-preserving synthesis: Maskgan for unpaired mr-ct translation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 56–65. Springer (2023)
18. Son, W., Park, C., Jeong, H.S., Song, Y.S., Lee, I.S.: Bone marrow edema in non-traumatic hip: high accuracy of dual-energy ct with water-hydroxyapatite decomposition imaging. European Radiology **30**, 2191–2198 (2020)
19. Yang, Q., Li, N., Zhao, Z., Fan, X., Chang, E.I.C., Xu, Y.: Mri cross-modality image-to-image translation. Scientific reports **10**(1), 3753 (2020)
20. 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)