DyeFreeNet: Deep Virtual Contrast CT Synthesis

Jingya Liu[1], Yingli Tian[1,⋆], A. Muhteşem Ağıldere[2], K. Murat Haberal[2], Mehmet Coşkun[2], Cihan Duzgol[3], and Oguz Akin[3]

[1] The City College of New York, New York, NY, USA 10031
[2] Baskent University, Ankara, Turkey 06810
[3] Memorial Sloan Kettering Cancer Center, New York, USA 10065

Abstract. To highlight structures such as blood vessels and tissues for clinical diagnosis, veins are often infused with contrast agents to obtain contrast-enhanced CT scans. In this paper, the use of a deep learning-based framework, DyeFreeNet, to generate virtual contrast abdominal and pelvic CT images based on the original non-contrast CT images is presented. First, to solve the overfitting issue for a deep learning-based method on small datasets, a pretrained model is obtained through a novel self-supervised feature learning network, whereby the network extracted intensity features from a large-scale, publicly available dataset without the use of annotations and classified four transformed intensity levels. Second, an enhanced high-resolution "primary learning generative adversarial network (GAN)" is then used to learn intensity variations between contrast and non-contrast CT images as well as retain high-resolution representations to yield virtual contrast CT images. Then, to reduce GAN training instability, an "intensity refinement GAN" using a novel cascade intensity refinement strategy is applied to obtain more detailed and accurate intensity variations to yield the final predicted virtual contrast CT images. The generated virtual contrast CTs by the proposed framework directly from non-contrast CTs are quite realistic with the virtual enhancement of the major arterial structures. To the best of our knowledge, this is the first work to synthesize virtual contrast-enhanced abdominal and pelvic CT images from non-contrast CT scans.

Keywords: Virtual Contrast CT · Image Synthesis · Self-supervised Learning · Deep Learning

1 Introduction

The use of contrast material is essential for highlighting blood vessels, organs, and other structures on diagnostic tests such as magnetic resonance imaging (MRI) and computed tomography (CT) [7–9]. However, contrast material may cause fatal allergic reactions or nephrotoxicity [1, 10]. This paper attempts to seek a dye-free solution by automatically generating virtual contrast-enhanced

⋆ Corresponding author. Email: ytian@ccny.cuny.edu

CTs directly from non-contrast CT images. There are existing studies based on image synthesis to assist the clinic diagnosis [6, 12, 15, 17, 19]. Recently, generative adversarial networks (GANs) [16] have shown to be promising for synthesizing medical images; for example, investigators have used GANs to synthesize 3D CT images from 2D X-rays with two parallel encoder-decoder networks [18], to synthesize MR images from non-contrast CT images, and to virtually stain specimens [2]. The synthesis of contrast-enhanced brain MR images from non-contrast or low contrast brain MR images has also been reported [3, 5].

In this paper, we focus on developing a new GAN-based framework to synthesize abdominal and pelvic contrast-enhanced CT images from non-contrast CT images, thereby virtually enhancing the arterial structures. Compared to synthesizing contrast-enhanced brain MR images, synthesizing contrast-enhanced CT images for the abdomen and pelvis is more challenging since they contain more feature and intensity variations. CT images are also susceptible to misregistration and there is a lack of multiparametric images to provide additional soft-tissue contrast. Additionally, abdominal and pelvic CT scans usually contain hundreds of CT slices with converging complex organs and soft tissue structures. To predict pixel intensity variations between synthesized contrast and the original non-contrast CTs accurately, the algorithm needs to obtain both local and global features. Lastly, with limited medical imaging data, the algorithm needs to account for overfitting during the training.

The contributions of this paper are summarized in the following three aspects. 1) **Virtual Contrast CT Synthesis.** To the best of our knowledge, this is the first work to synthesize virtual contrast-enhanced CT images from non-contrast abdominal and pelvic CT scans, which is more challenging than synthesizing contrast-enhanced brain MR images [3, 5]. 2) **Novel DyeFreeNet Framework.** This framework consists of the self-supervised learning network to obtain a pretrained model followed by high-resolution GANs to extract context features and predict intensity variations based on original non-contract CT images. We used a cascade intensity refinement strategy to train GANs in a progressive manner starting with key texture features and coarse intensity learning, followed by refining the intensity variations. 3) **Self-supervised Learning Pretrained Model.** To avoid overfitting and to allow the model to learn rich representative features, we first employed a novel self-supervised learning network to learn a pretrained model from a large-scale, publicly available dataset without the use of human annotations through classifying four intensity categories.

2 DyeFreeNet

The DyeFreeNet framework is proposed with the following two key aspects in mind. First, the virtual contrast CT image will be a contrast-enhanced version of the original non-contrast CT image whereby critical features of the original non-contrast CT image, such as the texture information of the body, organs, and soft tissues, will be preserved. Second, intensity variations in both local and global features of paired pre/post-contrast CT images will be accounted for in

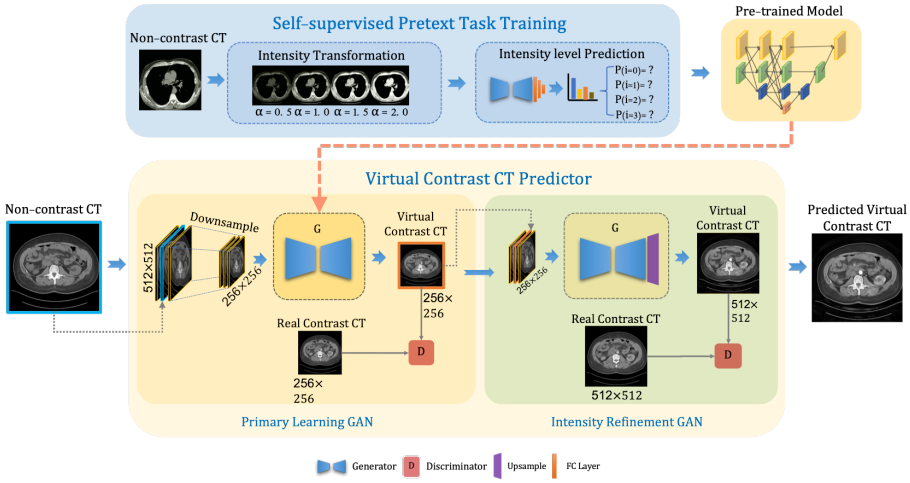

Fig. 1. The DyeFreeNet framework predicts virtual contrast CT images from non-contrast CT images by cascade intensity-learning high-resolution generative adversarial networks (GANs) combining the self-supervised feature learning schema. 1) A self-supervised learning pretrained model is trained by classifying four different intensity levels $(0.5, 1.0, 1.5, 2.0)$ transformed from non-contrast CT images that are available within a large public dataset. 2) A cascade training strategy is employed, whereby the "primary learning generative adversarial network (GAN)" learns the key texture features and coarse intensity variations from non-contrast CT images, and the "intensity refinement GAN" further refines contrast enhancement to yield the final predicted virtual contrast CT images.

network design and feature extraction. Fig. 1 shows the DyeFreeNet framework that consists of 1) A self-supervised pretrained model for rich feature extraction. 2) High-resolution intensity-learning GANs for preserving high-resolution features and virtual contrast CT generation using a cascade of intensity refinement training strategies.

2.1 Self-supervised Learning Pretrained Model

To speed up the training process and avoid overfitting for virtual contrast CT generation on relatively small dataset, self-supervised learning is proposed to extract rich intensity features from a large-scale, publicly available NLST dataset [11] with non-contrast CT without the use of data annotations and thereby obtain a pretrained model. As shown at Fig. 1, for each non-contrast CT image, intensity variances at four classes $[0, 1, 2, 3]$ are applied by adjusting the intensity coefficient $[0.5, 1.0, 1.5, 2.0, \textit{respectively}]$ to generate transformed CT images. A self-supervised intensity classification network employs the "high-resolution generator" from the virtual contrast CT predictor (see 2.2 below) as the backbone network for feature extraction. Extracted features are applied for the training

of a classifier to predict intensity level, using three fully connected layers with a softmax layer.

The cross-entropy loss function is shown in Eq.1:

$$loss(c_j|i) = -\frac{1}{K}\sum_{I=0}^{(K-1)} log(F(G(c_j, I)|i)), \tag{1}$$

where the input CT slice c_j is transformed into K levels of intensity I with the coefficient i. F indicates the classification network, and G is the intensity transformation model.

2.2 Virtual Contrast CT Predictor

The virtual contrast CT predictor consisted of cascade intensity refinement to generate high-resolution virtual contrast CT images.

Cascade Training Strategy. Due to the complex structures of abdomen and pelvis CT scans, cascade intensity refinement is split into two stages: 1) The "primary learning GAN" sketched the key texture features and coarse intensity variances. 2) The "intensity refinement GAN" focuses on refining and generating detailed contrast enhancement. The coarse-to-fine procedure utilizes both the spatial and temporal information of CT scans. The "primary learning GAN" takes three consecutive CT slices as input (mimics the RBG channels), and down-samples images to half of the original image size. It then generates the initial contrast CT image, with three continuous CT slices containing rich texture features but insufficient intensity variations serving as the input. The "intensity refinement GAN" takes the initial contrast CT generated by the "primary learning GAN" as the input and obtains more detailed intensity variations. Finally, an up-sampling layer is applied to yield the final predicted high-resolution virtual contrast CT images.

The GAN Architecture. High-resolution features are preserved using a proposed high-resolution encoder-decoder network similar to U-Net [13]. Inspired by High-Resolution Network (HRNet) [14], the encoder learns high-resolution features by back-propagating all layers with current convolutions block L concatenated through all previous convolutional blocks L_{i-1}. Meanwhile, the decoder network was constructed by skip connection, which gradually combines high-level features with low-level features. To preserve rich texture features from the original pre-contrast CT, the feature map from the last block was concatenated with the input features followed by two convolution layers for optimization.

As a "high-resolution generator" extracts the feature representatives, it is essential to map pre-contrast and contrast images in training accurately. We observed that traditional loss functions such as MSE and $L1$ losses easily oversmoothed the generated images. In comparison, GAN solves the issue by applying convolutional networks as discriminator distinguishing the real or generated images. By treating the virtual contrast CT as the result of a gradual approach from the pre-contrast CT to the virtual contrast CT, the virtual contrast CT

can be considered a regression task. The discriminator evaluated MSE and $BCEwithlogits$ losses for the following two sets of feature maps: 1) the feature map extracted from input CT concatenated with the real contrast CT, with the labels as ones; 2) the feature map extracted from input CT concatenated with the virtual CT, with the labels as zeros.

Objective Functions. The "high-resolution generator" learns the mapping between the pre-contrast image x and generated contrast image c to the real contrast image y. The generator G generates virtual contrast CT images, while the discriminator D is trained to distinguish the real and virtual contrast CT image. The objective of the proposed network DyeFreeNet (DFN) is shown as Eq. (2):

$$\mathcal{L}_{DFN}(G, D) = E_y[LogD(x, y)] + E_{x,c}[log(1 - D(G(x, c)))], \tag{2}$$

where G aims at minimizing the objective while D maximizes it. An additional MSE loss is appended with the objective to measure the distance between the virtual contrast image with the true image, shown as Eq. (3):

$$\mathcal{L}_{MSE}(G) = E_{x,y,c}||y - G(x, c)||_2^2. \tag{3}$$

To generate the virtual CT image similar to the real contrast CT image, MSE loss is conducted to optimize the intensity level close to real contrast CT gradually. The loss function of the discriminator is as Eq. (4):

$$\mathcal{L}_{DFN}(D) = D_{MSE}(G(x, c)) + D_{BCE}(G(x, c)). \tag{4}$$

Furthermore, perceptual loss [4] is applied for feature level comparison of generated and real virtual contrast CT images in between all the convolutional blocks. Therefore, the total objective of DyeFreeNet is:

$$G* = arg \min_G \max_D \mathcal{L}_{DFN} + \lambda \mathcal{L}_{MSE}(G) + \sum^{j} \ell_{feat}^{\phi,j}(y', y), \tag{5}$$

where λ adjusts the weight of MSE loss which set as 0.01, and y', y are the virtual and real features from the jth convolutional layer.

3 Experiments

Training and Validation Dataset. We assembled a retrospective CT dataset for examinations. All CT examinations were obtained using a dual-source multi-detector CT scanner. Patients were positioned supine on the table. Pre-contrast imaging of the abdomen was acquired from the dome of the liver to the iliac crest in an inspiratory breath hold by using a detector configuration of 192×0.6 mm, a tube current of 90 kVp, and a quality reference of 277 mAs. After intravenous injection of a 350 mg/ml non-ionic contrast agent (1.5 mL per kg of body weight at a flow rate of 4 ml/s), bolus tracking was started in the abdominal aorta at the level of the celiac trunk with a threshold of 100 HU. Scans were acquired

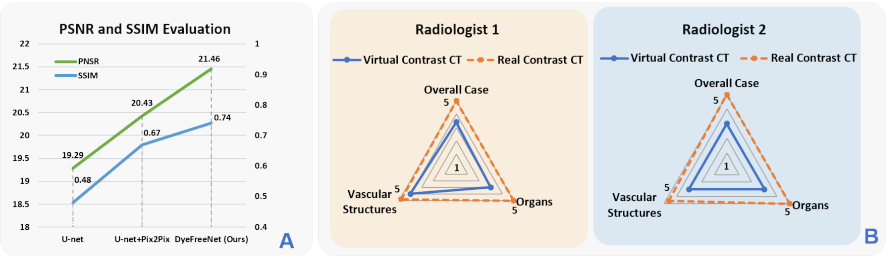

Fig. 2. A: The PSNR and SSIM results of the baseline models (U-Net and U-Net+Pix2Pix GAN) and the proposed DyeFreeNet on the comparison between virtual contrast CT images and the real contrast CT images. **B**: The evaluation scores from the assessments of two radiologists in three aspects: overall image quality, image quality of the organs, and the image quality of the vascular structures. Using a 5-point Likert scale from "1" = poor to "5" = excellent, the average evaluation score for virtual contrast CT images is "3" (acceptable). In particular, the image quality of the vascular structures was scored highest as the model was trained with early-stage arterial phase CT images.

using attenuation-based tube current modulation (CARE Dose 4D, Siemens). We focused on synthesizing the early-stage post-contrast CTs (vascular/arterial phase) from pre-contrast CTs. A total of 4,481 CT slices were used for training and validation while 489 CT slices were used for image quality evaluation.

Experimental Set Up and Parameter Settings.

Self-supervised Learning Pretrained Model. 41,589 CT slices of low-dose spiral CT scans were selected from the large public national lung screening trial (NLST) dataset [11] for the training of self-supervised pretrained model. The learning rate is set to $1e^{-6}$ and decreased by 0.1 after 5 epochs updated by Adam optimizer. The total training included 10 epochs with a batch size of 8.

Virtual Contrast CT Predictor. Three consecutive CT slices are fed as inputs to learn essential features in each image as well as between different CT slices. The "primary learning GAN" is initialized using the weights of the self-supervised pretrained model. The learning rate is set to $1e^{-5}$ and decreased by 0.1 after 20 epochs for 25 epochs training with a batch size of 4. The weight decay is $5e^{-4}$ with the Adam optimizer. The speed of virtual contrast image generation is 0.19 sec/slice on average with one GeForce GTX 1080 GPU using Pytorch 2.7.

The "intensity refinement GAN" is initialized using the weights of the trained primary learning model. The learning rate is $5e^{-5}$ and decreased by 0.1 after 17 epochs for 20 epochs training with a batch size of 4. The weight decay is $5e^{-4}$ using Adam optimizer. The speed of virtual contrast image generation is 0.24 sec/slice on a GeForce GTX 1080 GPU using Pytorch 2.7.

Quantitative Evaluation. Quantitative evaluation was performed between paired synthesized virtual contrast CT images and real contrast CT images that served as the ground truth for baseline models (U-Net and U-Net+Pix2Pix)

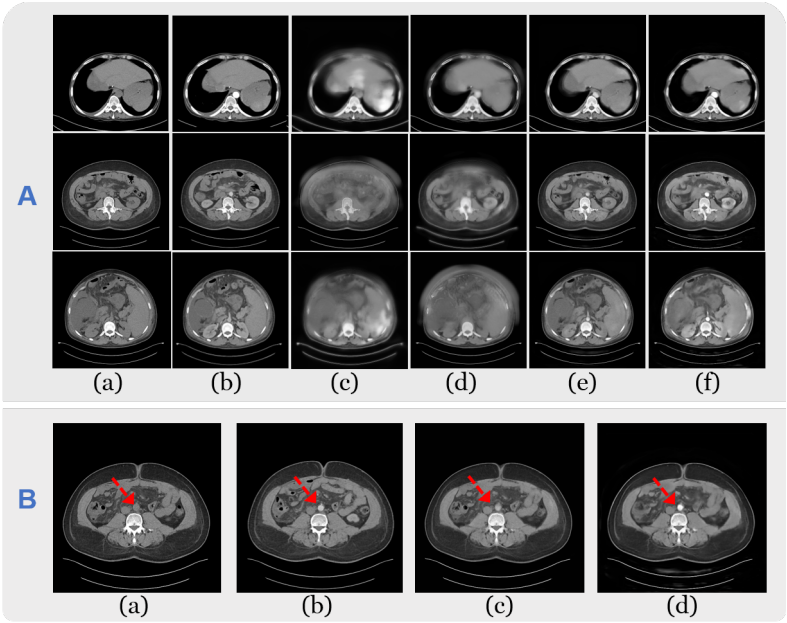

Fig. 3. A: The virtual contrast CT images generated from non-contrast CT images by the proposed DyeFreeNet compared with the baseline models of U-Net and U-Net+Pix2Pix GAN. a) Pre-contrast CT images. b) True contrast CT images as the ground truth. c) Virtual contrast CT images generated by U-Net. d) Virtual contrast CT images synthesized by U-Net+Pix2Pix GAN. e) Virtual contrast CT images synthesized by "primary learning GAN." f) Virtual contrast CT images predicted by "intensity refinement GAN." The illustration shows that our proposed framework could effectively synthesize high-resolution virtual contrast CT images similar to ground truth contrast CT images. **B**: The illustration of the contributions of the self-supervised learning-based pretrained model. (a) Pre-contrast CT image. (b) Real contrast CT image. (c) Virtual contrast CT image without the self-supervised pretrained model. (d) Virtual contrast CT with the self-supervised pretrained model. The red arrow indicates the intensity enhancement of the thoracic aorta region.

as well as our proposed DyeFreeNet. Voxel-wise difference and error assessment were conducted using Peak Signal to Noise Ratio (PSNR) while non-local structural similarity was assessed using Structural Similarity Index (SSIM). Fig. 2A shows that DyeFreeNet outperformed baseline models of U-Net and U-Net+Pix2Pix on PSNR (by 2.16 and 1.02, respectively), and also on SSIM (by 0.26 and 0.07, respectively).

Qualitative Evaluation by Radiologists. Blind reviews of paired pre-contrast CT images with real contrast CT images and paired pre-contrast CT images and virtual contrast CT images were conducted. Two radiologists independently assessed a total of 489 pairs of real and synthesized images on the following three aspects: overall image quality, image quality of the organs, and

image quality of the vascular structures. Qualitative scores were based on a 5-point Likert scale, with scores ranging from "1" = poor, "2" = sub-optimal, "3" = acceptable, "4" = good, and "5" = excellent. Fig. 2B shows that the radiologists gave an average score of "3" (acceptable) for overall image quality, image quality of the organs, and image quality of the vascular structures for the virtual contrast images compared with an average score of "5" (excellent) for the real contrast images. The average score for the image quality of the vascular structures was slightly higher than the overall image quality and image quality of the organs. In contrast, the average score was slightly lower for the image quality of the organs, likely because the virtual contrast images were generated from the network trained with vascular (arterial) phase images.

Comparison with Baseline Models. Additional results are illustrated in Fig. 3A. Pre-contrast and real contrast CT images are shown in Fig. 3A(a) and (b), respectively; enhanced regions on the real contrast CT images depict the arterial structures. Virtual contrast CT images generated by U-Net, U-Net+Pix2Pix GAN, "primary learning GAN" of DyeFreeNet, and "intensity refinement GAN" of DyeFreeNet are shown in Fig. 3A(c-f). Although U-Net partially learned the intensity variations, the virtual contrast CT is very blurry and includes artifacts. While traditional MSE or $L1$ loss functions can be applied, they easily oversmooth the predicted image which is problematic as high-resolution images are required for diagnosis. By using the "primary learning GAN," texture features are successfully preserved from the pre-contrast images. However, although the resolution is increased, the intensity variance learning is decreased. Using a pretrained model with "primary learning GAN" results in better feature extraction but the intensity variations still need to be improved. With the "intensity refinement GAN," the DyeFreeNet accurately enhances the vascular structures. The intensity variations and texture features are both learned and preserved in this framework with contributions by the self-supervised learning pre-trained model and the cascade framework.

Comparison with Self-supervised Pre-trained Model. Fig. 3B illustrates the results with and without using the self-supervised intensity pretrained model. Pre-contrast and real post-contrast CT images are shown in Fig. 3B(a) and (b). Although the thoracic aorta region (red arrow) is enhanced without using the pretrained model as shown in Fig. 3B(c), with rich feature extraction from the pretrained model, the intensity variations are significantly enhanced as shown in Fig. 3B(d).

Remaining Challenges and Future Work In this paper, the performance of the model is evaluated at the arterial stage. Our future work will seek to extend the DyeFreeNet network for multi-stage virtual contrast generation (i.e., portal and delayed phases). Its potential limitation is that the misalignment between pre-contrast CT and contrast CT (as the training data) might be more significant than the arterial phase, resulting in generating the artifacts that affect the enhancement accuracy. It is essential to tackle the misalignment issue. Furthermore, to validate the possibility of clinical practice, a downstream task

assessment will be developed in future work, such as nodule detection, blood vessel segmentation, and organ segmentation.

4 Conclusion

We developed a self-supervised intensity feature learning-based framework, DyeFreeNet, to automatically generate virtual contrast-enhanced CT images from non-contrast CT images. The rich features extracted by the self-supervised pre-trained model and a coarse-to-fine cascade intensity refinement training schema significantly contributed to high-resolution contrast CT image synthesis. The promising results show high potential to generate virtual contrast CTs for clinic diagnosis.

Acknowledgements

This material is based upon work supported by the National Science Foundation under award number IIS-1400802 and Memorial Sloan Kettering Cancer Center Support Grant/Core Grant P30 CA008748.

References

1. Andreucci, M., Solomon, R., Tasanarong, A.: Side effects of radiographic contrast media: pathogenesis, risk factors, and prevention. BioMed research international **2014** (2014)
2. Bayramoglu, N., Kaakinen, M., Eklund, L., Heikkila, J.: Towards virtual h&e staining of hyperspectral lung histology images using conditional generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 64–71 (2017)
3. Gong, E., Pauly, J.M., Wintermark, M., Zaharchuk, G.: Deep learning enables reduced gadolinium dose for contrast-enhanced brain mri. Journal of Magnetic Resonance Imaging **48**(2), 330–340 (2018)
4. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European conference on computer vision. pp. 694–711. Springer (2016)
5. Kleesiek, J., Morshuis, J.N., Isensee, F., Deike-Hofmann, K., Paech, D., Kickingereder, P., Köthe, U., Rother, C., Forsting, M., Wick, W., et al.: Can virtual contrast enhancement in brain mri replace gadolinium?: A feasibility study. Investigative radiology (2019)
6. Li, Z., Wang, Y., Yu, J.: Brain tumor segmentation using an adversarial network. In: International MICCAI Brainlesion Workshop. pp. 123–132. Springer (2017)
7. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical image analysis **42**, 60–88 (2017)
8. Liu, J., Li, M., Wang, J., Wu, F., Liu, T., Pan, Y.: A survey of mri-based brain tumor segmentation methods. Tsinghua Science and Technology **19**(6), 578–595 (2014)

9. Liu, J., Cao, L., Akin, O., Tian, Y.: 3dfpn-hs^2: 3d feature pyramid network based high sensitivity and specificity pulmonary nodule detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 513–521. Springer (2019)
10. Mentzel, H.J., Blume, J., Malich, A., Fitzek, C., Reichenbach, J.R., Kaiser, W.A.: Cortical blindness after contrast-enhanced ct: complication in a patient with diabetes insipidus. American journal of neuroradiology **24**(6), 1114–1116 (2003)
11. National Lung Screening Trial Research, T.: The national lung screening trial: overview and study design. Radiology **258**(1), 243–253 (2011), [dataset]
12. Ren, J., Hacihaliloglu, I., Singer, E.A., Foran, D.J., Qi, X.: Adversarial domain adaptation for classification of prostate histopathology whole-slide images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 201–209. Springer (2018)
13. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
14. Sun, K., Xiao, B., Liu, D., Wang, J.: Deep high-resolution representation learning for human pose estimation. arXiv preprint arXiv:1902.09212 (2019)
15. Wei, W., Poirion, E., Bodini, B., Durrleman, S., Ayache, N., Stankoff, B., Colliot, O.: Learning myelin content in multiple sclerosis from multimodal mri through adversarial training. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 514–522. Springer (2018)
16. Yi, X., Walia, E., Babyn, P.: Generative adversarial network in medical imaging: A review. Medical image analysis p. 101552 (2019)
17. Ying, X., Guo, H., Ma, K., Wu, J., Weng, Z., Zheng, Y.: X2ct-gan: Reconstructing ct from biplanar x-rays with generative adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 10619–10628 (2019)
18. Ying, X., Guo, H., Ma, K., Wu, J., Weng, Z., Zheng, Y.: X2ct-gan: Reconstructing ct from biplanar x-rays with generative adversarial networks. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)
19. Zhao, H.Y., Liu, S., He, J., Pan, C.C., Li, H., Zhou, Z.Y., Ding, Y., Huo, D., Hu, Y.: Synthesis and application of strawberry-like fe3o4-au nanoparticles as ct-mr dual-modality contrast agents in accurate detection of the progressive liver disease. Biomaterials **51**, 194–207 (2015)