공학석사 학위논문

흉부 CT 영상에서의 재구성 커널 변환을 위한 비지도 적대적 생성 신경망 네트워크 개발

Development of unsupervised generative adversarial network for reconstruction kernel conversion in chest CT imaging

울산대학교대학원

의과학과

최창용

흉부 CT 영상에서의 재구성 커널 변환을 위한 비지도 적대적 생성 신경망 네트워크 개발

지도교수 김남국

이 논문을 공학석사 학위논문으로 제출함

2024년 2월

울산대학교대학원

의과학과

최창용

최창용의 공학석사 학위논문을 인준함

심사위원 이상민 인

심사위원 김남국 인

심사위원 이준구 인

울산대학교대학원

2024년 2월

감사의 글

가장 먼저 26년이라는 시간 동안 나를 위해 헌신하며 희생하신 아버지와 어머니께 감사하다는 말씀드립니다. 제가 어떤 길을 선택하든 그저 존중하며 넘어지지 않도록 등을 받쳐주어 나아가는 길에 두려움이 없는 천군만마(千軍萬馬)와 같은 힘을 받을 수 있었습니다. 더불어 함께 살아오면서 말다툼도 하고 내가 못되게 대한 적도 많지만 그럼에도 항상 아낌없이 챙겨주고 부족함을 지적해주는 둘도 없는 나의 친형에게도 고맙다는 말을 전합니다. 온화하고 화목한 가정의 분위기 속에서 학문과 인성의 정도(正道)를 걸을 수 있게 해 주었음에 감사합니다.

석사 과정을 거치는 동안 사수로서 내가 나아가야 할 방향을 제시해주었던 경성구, 정지헌 선생님께 감사하다는 말을 전합니다. 두 분 덕에 석사 기간 동안 방황하지 않고 즐겁고 보람찬 연구를 할 수 있었습니다. 또한 나에게 많은 시간을 할애하여 부족한 수학적 지식을 주고자 많은 것을 가르쳐주었던 김인환 선생님께 감사하다는 말을 전합니다. 서로 어렵거나 힘든 일이 있을 때 들어주며 공감하고 의지할 수 있었던 나의 동기 김지영, 박강길 선생님께도 감사하다는 말을 전합니다. 그 외에 함께 연구를 했던 김기덕, 남유진, 이소영, 조경진 선생님과 짧은 시간이었지만 많은 도움을 줬던 부사수 이상윤 선생님, 연구실 밖에서 나와 함께 어울려주었던 김영제, 김지환, 임지섭 선생님께 감사하다는 말을 전합니다. 모든 분들 덕에 행복한 연구실 생활을 할 수 있었고 나의 인생에는 인복(人福)이라는 행운이 항상 따르고 있다는 것을 다시 한번 상기할 수 있는 기회가 되었습니다. 마지막으로 올바르고 의미 있는 연구를 하고 좋은 결과를 맺을 수 있도록 이끌어주고 조언해주신 김남국 교수님, 이준구 교수님 그리고 이상민 교수님께 감사하다는 말씀드립니다.

Abstract

Computed tomography (CT) image is one of the diagnostic imaging widely used in the medical field. CT image is reconstructed from sinogram, which is the 2D array data containing the projections, using convolution kernel through back projection. At this point, the kernel differs depending on which anatomical structure is evaluated in qualitative evaluation. Also, quantitative evaluation is crucial as well as qualitative evaluation and affects the choice of kernel. However, there are two problems. First, sinogram has large capacity and storage space is limited, so CT image is usually reconstructed with only one specific kernel for evaluation and sinogram is removed in a week. Second, patients should be scanned and exposed radiation once again. Recently, many researchers have proposed image-to-image translation methods using generative adversarial networks (GANs) for CT kernel conversion. Nevertheless, preserving anatomical structure including fine details, e.g., airway and blood vessel, while transferring the style of the target kernel is still challenging when CT image is translated from the source kernel to the target kernel. In this study, kernel conversion GAN (KCGAN) is proposed to alleviate these problems with perceptual guidance and showed robust and efficient performance in kernel conversion. Perceptual guidance is a type of discriminator regularization method using feature map of generator to learn semantic representation better. For content and style features, cosine similarity content loss and contrastive style loss are defined between the feature map of generator and semantic label map of discriminator, respectively. KCGAN can preserve the fine-grained anatomical structure of the source domain and transfer the style of the target domain, simultaneously. In addition, this method can be easily applied with only changing the discriminator architecture and without utilizing any additional learnable or pre-trained networks. Experimental results showed that this method outperformed existing GAN-based methods in most direction of kernel conversion among three kernels.

Contents

Contents of Tables

Contents of Figures

Introduction

Computed Tomography (CT) image is now one of the diagnostic imaging widely used in medical field and has recently been used for screening for disease. CT image is acquired through back-projection from sinogram, which is 2D array projection data collected from rotating X-ray tube and detectors. Before back-projection, convolution kernel is applied as a kind of technical parameter and filter that changes frequency of the reconstructed CT image. Depending on what kernel being used for CT image reconstruction, there is a trade-off between the spatial resolution and noise, so it affects texture quantitative values [1–3]. The sharp kernel makes CT image have high spatial resolution and noise, and this can help to screen abnormality in bone or lung. On the other side, the soft kernel makes CT image have low spatial resolution and noise, and this can help to screen abnormality in soft tissue or mediastinum. Likewise, CT image needs to be reconstructed with different kernels depending on which anatomical structure is being evaluated. Furthermore, even if the same anatomical structure is being evaluated, the kernel being used is different depending on whether qualitative or quantitative evaluation is performed. For instance, in chest CT image to analyze chronic obstructive pulmonary disease (COPD), it is reconstructed with soft kernel for quantitative evaluation and with sharp kernel for qualitative evaluation. For these reasons, CT images reconstructed with various kernels are sometimes necessary for more accurate diagnosis.

However, there are limitations in reconstructing CT images with different kernels. First, sinogram has large capacity, so CT image is reconstructed with only one specific kernel for evaluation and sinogram is usually removed in a week. Even though CT image is reconstructed with all kinds of kernels, storage space is limited. Therefore, medical doctors have had difficulty evaluating qualitatively or quantitatively without CT images reconstructed with other kernels. Particularly, this limitation reveals on retrospective or longitudinal studies because they cannot control kernels which are technical parameters [2]. Consequently, patients should be scanned again to acquire new CT images, then exposed to unnecessary radiation.

Many studies have proposed image-to-image translation (I2IT) methods [4] using convolutional neural network (CNN)-based CT kernel conversion [2,3,5,6]. However, there is a disadvantage of CNN that must be a paired dataset for training. Recently, instead of CNN, Generative adversarial network (GAN) [7]-based CT kernel conversion [8,9] have been proposed due to unsupervised manner and the powerful generation ability. But when using unsupervised I2I (UI2IT) methods, preserving anatomical structure including fine details, e.g., airway and blood vessel, while transferring the style of the target image is still challenging when translated from the source image to the target image, especially in medical domain [10]. This can cause large pixel intensity variation and be critical for analyzing quantitative evaluation. In this study, kernel conversion GAN (KCGAN) is proposed to alleviate these limitations using chest CT images with perceptual guidance which improves performance of multi-domain UI2IT in kernel conversion. Perceptual guidance is proposed as a new type of discriminator regularization method using feature map of generator to learn semantic representation better. For content and style features, cosine similarity content loss and contrastive style loss are defined between the feature map of generator and semantic label map of discriminator, respectively.

The contributions of KCGAN with perceptual guidance are as follows:

- KCGAN preserves the coarse-to-fine anatomical structure of the source image.
- Perceptual guidance can be easily applied with only changing the discriminator architecture and without utilizing any additional learnable or pre-trained networks and encoding process.
- Experimental results showed that this method outperformed existing GAN-based methods in CT kernel conversion.

Materials and Method

Datasets

This retrospective study was approved by the institutional review board of Asan Medical Center and written informed consent was waived. A total of 170 patients scanned CT images and consisted of unpaired 150 patients for training and paired 20 patients for test. CT images were obtained using Somatom Definition Edge, AS, AS+ and Flash; Siemens Healthineers, Forchheim, Germany.

For train datasets, from January 2015 to July 2021, unpaired non-contrast chest CT images reconstructed with B30f (soft), B50f (standard) and B70f (sharp) kernels were collected from 50 patients (16760 slices; 37 men and 13 women; mean age, 61.7 ± 13.5 [SD] years), 50 patients (16339 slices; 24 men and 26 women; mean age, 66.7 ± 13.1 [SD] years) and 50 patients (17042 slices; 25 men and 25 women; mean age, 62.6 ± 12.3 [SD] years), respectively.

For test datasets, from April 2017 to July 2021, paired non-contrast chest CT images reconstructed with same kernels as trainset were collected from 20 patients (6897 slices; 15 men and 5 women; mean age, 67.1 ± 7.4 [SD] years) for quantitative and qualitative evaluations. Other CT acquisition parameters are shown in Table 1.

Table 1. CT acquisition parameters of dataset according to type of kernel.

	Kernel	Patients	Slices	Age (year)	Sex (M:F)	Slice Thickness	kVp	mAs
Train	B30f	50	16760	61.7 ± 13.5	37:13	1.0	120	120
	B50f	50	16339	66.7 ± 13.1	24:26	1.0	120	120
	B70f	50	19469	62.6 ± 12.3	32:23	1.0	120	120
	Kernel	Patients	Slices	Age (year)	Sex (M:F)	Slice Thickness	kVp	mAs
Test	B30f	20	6897	67.1 ± 7.4	15:5	1.0	120	120
	B50f	20	6897	67.1 ± 7.4	15:5	1.0	120	120
	B70f	20	6897	67.1 ± 7.4	15:5	1.0	120	120

Multi-domain Image-to-image Translation

Image-to-image translation (I2IT) aims to mapping from a source domain to a target domain. Since GAN came out in 2014, I2IT made tremendous progress. Pix2Pix [11] is representative supervised I2IT model using conditional input to translate an image. However, Pix2Pix needs paired train datasets, so researchers have proposed unsupervised I2IT (UNI2IT) [12–17] for unpaired datasets. CycleGAN [12] was proposed as UNI2IT model that uses two generators, one translates from the source domain to the target domain, and the other translates from the translated domain back to the source domain. CycleGAN optimizes these two generators to reduce the difference between the source domain and the translated source domain using cycle-consistency loss [12]. Although CycleGAN breaks paired constraint, it suffers from the limitation which cannot translate from one domain to multiple domains at once. As multi-domain UNI2IT model, StarGAN [13] was proposed and it could translate the source domain to multiple domains by embedding various attributes as mask vectors.

I2IT task has also been utilized in medical domain. Many researchers have proposed for CT kernel conversion [2,3,5,6], however, they used convolutional neural networks (CNN) that requires paired datasets. Gravina, M., et al. [8] used CycleGAN for unpaired manner, but it was still for only two domains. Therefore, they should train multiple networks for each kernel conversion. Switchable CycleGAN [9] solved this two domains limitation and proposed a continuous kernel conversion using adaptive instance normalization (AdaIN) [18]. This method has advantages of interpolating images from the source domain to the target domain and translating various kernels between the two domains but could not preserve fine-grained anatomical structure perfectly.

In this study, StarGAN [13] which is one of multi-domain UNI2IT models was exploited as baseline model to translate kernels to all directions at once. StarGAN consists of generator that has an encoder and a decoder and discriminator that performs multi-task for adversarial classification and attributes classification. The losses of discriminator and generator are as follows:

$$\mathcal{L}_D = -\mathcal{L}_{adv} + \lambda_{cls}\mathcal{L}_{cls}^{r}, \tag{1}$$

$$\mathcal{L}_G = \mathcal{L}_{adv} + \lambda_{cls}\mathcal{L}_{cls}^{f} + \lambda_{cyc}\mathcal{L}_{cyc}, \tag{2}$$

where $\mathcal{L}_D$ and $\mathcal{L}_G$ are the discriminator and generator losses, respectively. $\mathcal{L}_{adv}$ is the adversarial loss for binary classification between real and generated images. $\mathcal{L}_{cls}^{r}$ and $\mathcal{L}_{cls}^{f}$ are the domain attributes classification losses for a real and generated image, respectively. $\mathcal{L}_{cyc}$ is the cycle consistency loss. λ_{cls} and λ_{cyc} are the hyperparameters that weight the importance of the domain attributes classification losses and the cycle consistency loss. These two hyperparameters are set 1 and 10 like configuration of StarGAN, respectively.

Regularization for Discriminator in GANs

Several studies have proposed regularization methods for the balance between the discriminator and the generator such as differentiable augmentation [19], gradient penalty [20], spectral normalization [21] and generator-guided discriminator regularization [22].

DiffAugment [19] for data efficient training applies the augmentation to both the real image and the generated image when discriminator is updated and to the generated image when generator is updated, because augmentation function should be differentiable. DiffAugment is efficient when the number of training data lacks, so this method wasn't used in this study. Gradient penalty [20] is a soft version of the Lipschitz constraint, which gives gradient norm penalty instead of weight clipping [23]. This can be GAN training more stable and has been used in recent state-of-the-art GAN including StarGAN [13].

Spectral normalization [21] uses Lipshitz constant, which is the only hyper parameter and doesn't require tuning, to regularize discriminator instead of batch normalization [24]. It also showed tremendous increased performance of GAN and quality of the output image likewise other regularization methods. Recently, Generator-guided discriminator regularization (GGDR) [22] was proposed as a new type of discriminator regularization method that discriminator can learn semantic representation and extract semantic label map by comparing with intermediate feature map from generator for unconditional generation. This method improves fidelity as much as conditional GAN [25–27] without any ground-truth semantic label maps. Our proposed method is directly inspired by GGDR.

Cosine Similarity Content Loss

In GGDR, cosine distance loss was applied between the intermediate feature map and the semantic label map. This is for convenience of scaling within a specified range because of balance with adversarial loss. Cosine distance loss would be helpful for discriminator to learn overall semantic content representation of translated image in terms of I2IT, so we define this loss as cosine similarity content loss (CCL) and this loss function follows as:

$$\mathcal{L}_{ccl} = \mathbb{E}_{x\sim p(x)}\left[1 - \frac{F(G^{\ell}(x)) \cdot G^{\ell}(x)}{\|F(G^{\ell}(x))\|_2 \cdot \|G^{\ell}(x)\|_2}\right], \tag{3}$$

where x is input images from the source domain. F is the decoder of discriminator and $G^{\ell}(\cdot)$ is the ℓth decoder layer of generator. The resolution of semantic label map from F is the same with that of the intermediate feature map from G^{ℓ}.

Contrastive Style Loss

In I2IT, it is important to learn style as well as content, so CCL may be not sufficient to segment a detailed semantic label map. Recently, contrastive learning has been proved to be effective in learning semantic label maps for semantic segmentation and object detection [28,29]. Also, it has shown superior performance that patch-wise contrastive loss can transfer the style of the target image for I2IT [16], even as a substitute of cycle-consistency loss. Unlike image generation task, I2IT task needs generator which consists of an encoder, residual block [30] and a decoder, this structure design is a core for translating the style of the source image. Through the decoder, source image is getting closer to translated image and the feature map in decoder layer might contain the style of the target image. For these reasons, the intermediate feature map from the decoder of the generator might be effective for discriminator to learn fine-grained semantic style representation. We use PatchNCE loss [16] and define this loss as contrastive style loss (CSL), and this loss function follows as:

$$\mathcal{L}_{csl} = \mathbb{E}_v \left[-\log \frac{\exp(v \cdot v^+/\tau)}{\exp(v \cdot v^+/\tau) + \sum_{n=1}^{N} \exp(v \cdot v_n^-/\tau)} \right], \quad (4)$$

where v is a vector which are mapped with query patch from the semantic label map of discriminator decoder and v^+, v_n^- are vectors which are mapped with positive and n-th negative patches from the intermediate feature map of generator decoder, respectively. τ is the hyperparameter, but we set τ = 0.07 like configuration of CUT. $\mathcal{L}_{csl}$ performs (N+1)-way classification that the positive patch is made up to associate with the query patch more than the negative patches.

Perceptual Guidance

CCL plays a role to inform coarse semantic content representation and CSL plays a role to inform fine-grained semantic style representation. Here, we propose perceptual guidance which is combined with CCL and CSL so that the discriminator can learn both content and style feature from the generator to improve performance of multi-domain UNI2IT. Perceptual guidance is inspired from perceptual loss [31,32] which consists of content loss and style loss in feature domain through VGG model [33]. However, unlike perceptual loss, perceptual guidance only needs two feature maps from the discriminator and the generator themselves, so it doesn't require any additional learnable or pre-trained feature extractor networks and encoding process. It is defined as follows:

$$\mathcal{L}_{pg} = \lambda_{ccl}\mathcal{L}_{ccl} + \lambda_{csl}\mathcal{L}_{csl}, \quad (5)$$

where λ_{ccl} and λ_{csl} are the hyperparameters for balance between $\mathcal{L}_{ccl}$ and $\mathcal{L}_{csl}$. These two hyperparameters are respectively set 1 and 5 and will be shown in experiments and results section for balancing them. $\mathcal{L}_{pg}$ is added when the discriminator updates, so the discriminator conducts multi-task learning [34]—real and fake classification, domain attributes classification and semantic segmentation. The total loss functions for the discriminator and the generator follows as:

$$\mathcal{L}_D = -\mathcal{L}_{adv} + \lambda_{cls}\mathcal{L}_{cls}^r + \mathcal{L}_{pg}, \quad (6)$$

$$\mathcal{L}_G = \mathcal{L}_{adv} + \lambda_{cls}\mathcal{L}_{cls}^f + \lambda_{cyc}\mathcal{L}_{cyc}. \quad (7)$$

The overall framework of perceptual guidance is shown in Figure 1.

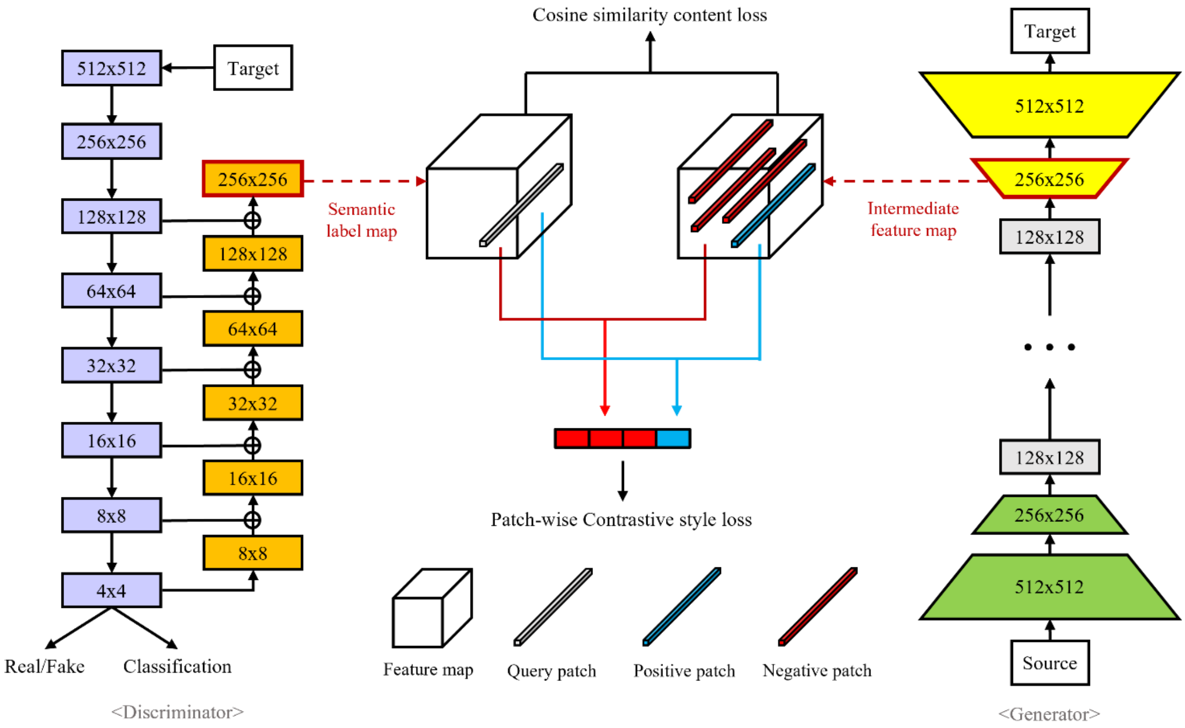

Figure 1. Overall framework of perceptual guidance. Perceptual guidance learns coarse semantic content representation and fine-grained semantic style representation through cosine similarity content loss and patch-wise contrastive style loss between intermediate feature map from decoder of generator and semantic label map from decoder of discriminator.

Implementation Details

The resolution of CT images was not resized and maintained the size 512 × 512. The intensities of CT images were normalized from their full range of Hounsfield units (HU) (-1024–3071) to (-1–1) as pre-processing. As optimization for the discriminator and the generator, Adam [35] was used with $\beta_1 = 0.5$ and $\beta_2 = 0.999$ and the learning rate is 1e-4. Gradient penalty [20] is used with critics as 5, where critics is the number of updates for the discriminator per each update of the generator. The intermediate feature map from the decoder of the generator was extracted with the size 256 × 256. The number of patches was 64. All experiments including other GAN-based UNI2IT models were conducted using single NVIDIA GeForce RTX3090 24GB GPU for 16 epochs with batch size 2.

Architecture Improvements

We selected StarGAN [13] as a baseline model which showed plausible performance in multi-domain translation, however, other GAN-based UNI2IT models [14,16,17] has come out superior to StarGAN. Nevertheless, by applying perceptual guidance, adapting spectral normalization [21] to the discriminator and pixelshuffle [36]—up-sampling method for high quality image—to the generator, we thought that it can still have possibility to increase performance of StarGAN and catch-up other GAN-based UNI2IT models. Architectural ablation studies were implemented empirically for the best quality results. At the up-sampling layers in decoder block of the generator, 4 × 4 transposed convolution was applied but it caused degradation of visual quality results because of checkerboard artifact [37]. Using 3 × 3 convolution with 2 × 2 pixelshuffle could prevent the checkerboard artifact and showed better visual quality results than the transposed convolution. Meanwhile, the discriminator followed U-Net [38] architecture with skip connection to extract the semantic label map through the intermediate feature map from the generator. It consisted of encoder block which classifies real or fake images and domain attributes of images through N down-sampling layers, and decoder block which extracts the semantic label map through $N-1$ up-sampling layers by matching the intermediate feature map of the generator. Since the semantic label map should have the same size with the intermediate feature map, we notify that the decoder block has $N-1$ up-sampling layers. For skip connections, each feature map from the down-sampling and the up-sampling layers was concatenated together, then 1 × 1 convolution with 2 × 2 pixelshuffle was applied. Additionally, spectral normalization and leakyReLU activation function were applied in all layers of the discriminator instead of instance normalization and ReLU activation function. Through this application of architectural changes with perceptual guidance, we propose kernel conversion GAN (KCGAN).

Experiments and Results

We compared KCGAN with two-domain UNI2IT models such as CycleGAN [12], CUT [16] and DCLGAN [17] and multi-domain UNI2IT models such as StarGAN [13] and AttGAN [14]. In this section, qualitative and quantitative results were evaluated 6 directions—B30f to B50f, B30f to B70f, B50f to B30f, B50f to B70f, B70f to B30f and B70f to B50f—about kernel conversion. Also, visualization of feature maps from the discriminator and the generator was conducted for better comprehension. Lastly, ablation studies about the balance of the hyperparameters λ_{ccl} and λ_{csl}, separate usage of CCL and CSL, usage of encoder feature map from the generator for CCL, effect of perceptual guidance and computational cost were conducted to prove efficiency of our proposed method.

Comparison with GAN-based Unsupervised Image-to-image Translation Models

We showed the qualitative results of kernel conversion about 6 directions using GAN-based UNI2IT models including KCGAN. For qualitative result visualization, window width and level were set 1500 and -700, respectively. As shown in Figure 2, for each kernel conversion, figure is consisted of whole image and airway zoomed image in first row and second row, respectively. While CycleGAN and DCLGAN could not preserve the overall sparse anatomical structure when compared to the target image and DCLGAN showed worst performance, the other models showed plausible qualitative results.

Nevertheless, if the translated images are zoomed in and looked deeply into the airway and the vessel, they also could not preserve fine-grained anatomical structure in the translation about most direction. For better understanding the structural differences between the target image and the translated image, we visualized difference maps in Figure 3. More details for each kernel conversion are as follows:

1) The translation from sharp kernels to soft kernels (B50f to B30f, B70f to B30f and B70 to B50f): CT images translated from sharp kernels to soft kernels should look blur and have less noise. It is shown that AttGAN could not translate blurriness as much as the target kernel and created new vessels that did not exist before. As seen in B50f to B30f and B70f to B50f, the translated image generated by CycleGAN created artifact around soft tissues and lung, and the translated image generated by DCLGAN removed the original information or created new information that did not exist before. Additionally, they could not preserve the coarse anatomical structure as well as the fine-grained details. In AttGAN, CUT and StarGAN, it is shown that the thickness of airway wall is different or disconnected. In addition, the translated images generated by StarGAN have more noise than the target images and too much blurriness. On the other hand, KCGAN showed great translation performance preserving the anatomical structure including airway and vessel. These results showed that our proposed method could preserve the anatomical structure better than the other models.

2) The translation from soft kernels to sharp kernels (B30f to B50f, B30f to B70f and B50f to B70f): CT images translated from soft kernels to sharp kernels should look clear and have a lot of noise. Like the translation from sharp kernels to soft kernels, the translated images generated by CycleGAN have disconnected airway wall and the translated images generated by CUT have different thickness of airway wall. DCLGAN could not preserve the anatomical structure badly. As seen in B30f to B50f, the translated image generated by AttGAN could not express the blurriness and noise as much as the target image and the translated image generated by StarGAN showed hatch pattern artifact, so it degraded the qualitative result. In B30f to B70f, CycleGAN and CUT created some artifacts that looks like vessels. In B50f to B70f, all results showed plausible qualitative results that translates blurriness and noise level. However, they still have weakness in preserving the fine-grained details. StarGAN showed overall poor quality of the translated images. Although our model is based on vanilla StarGAN, it is shown that our proposed method could improve performance much better.

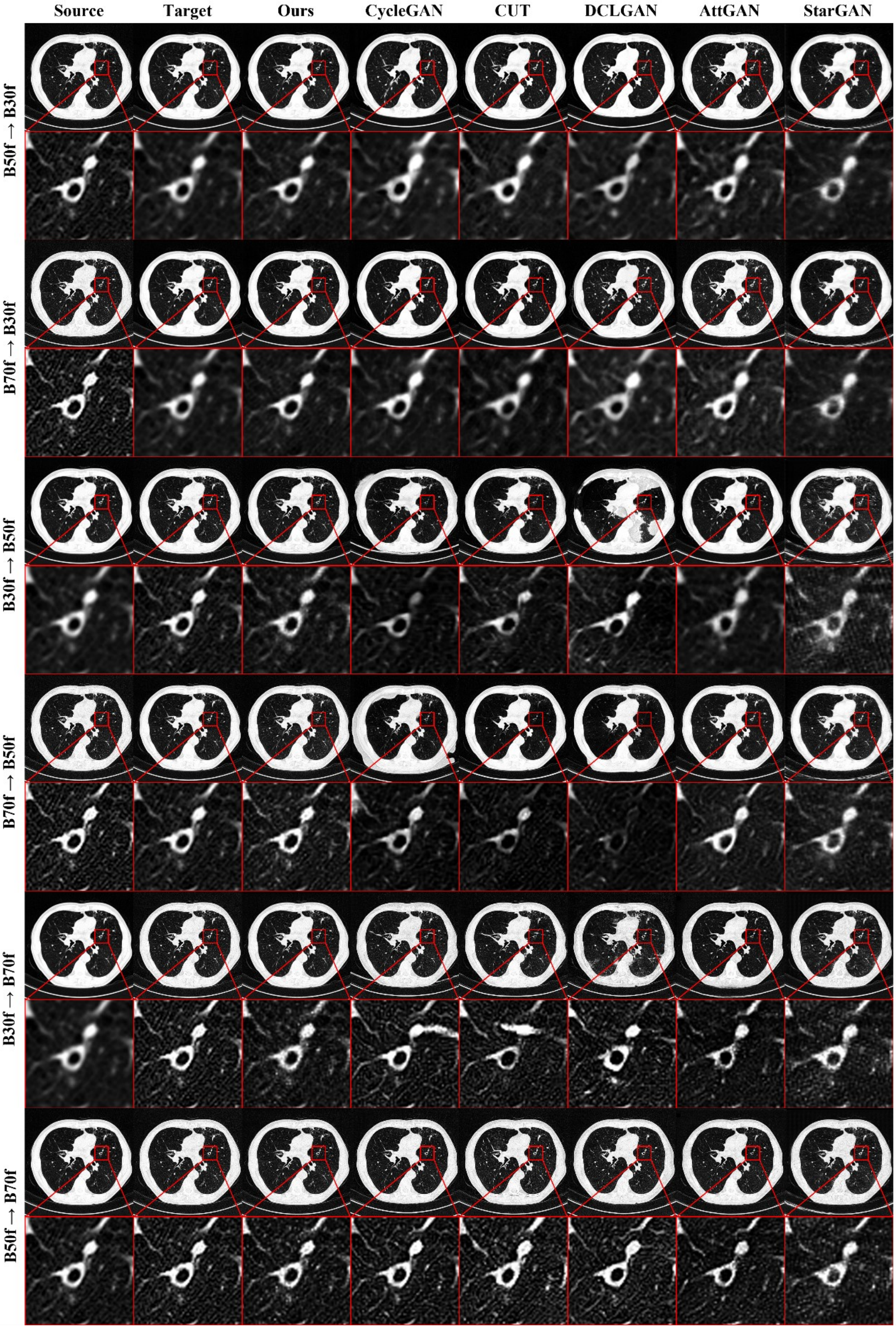

Figure 2. The qualitative results of kernel conversion about 6 directions. For each kernel conversion, first row is a whole chest CT image, and second row is a zoomed image which points out airway and vessels.

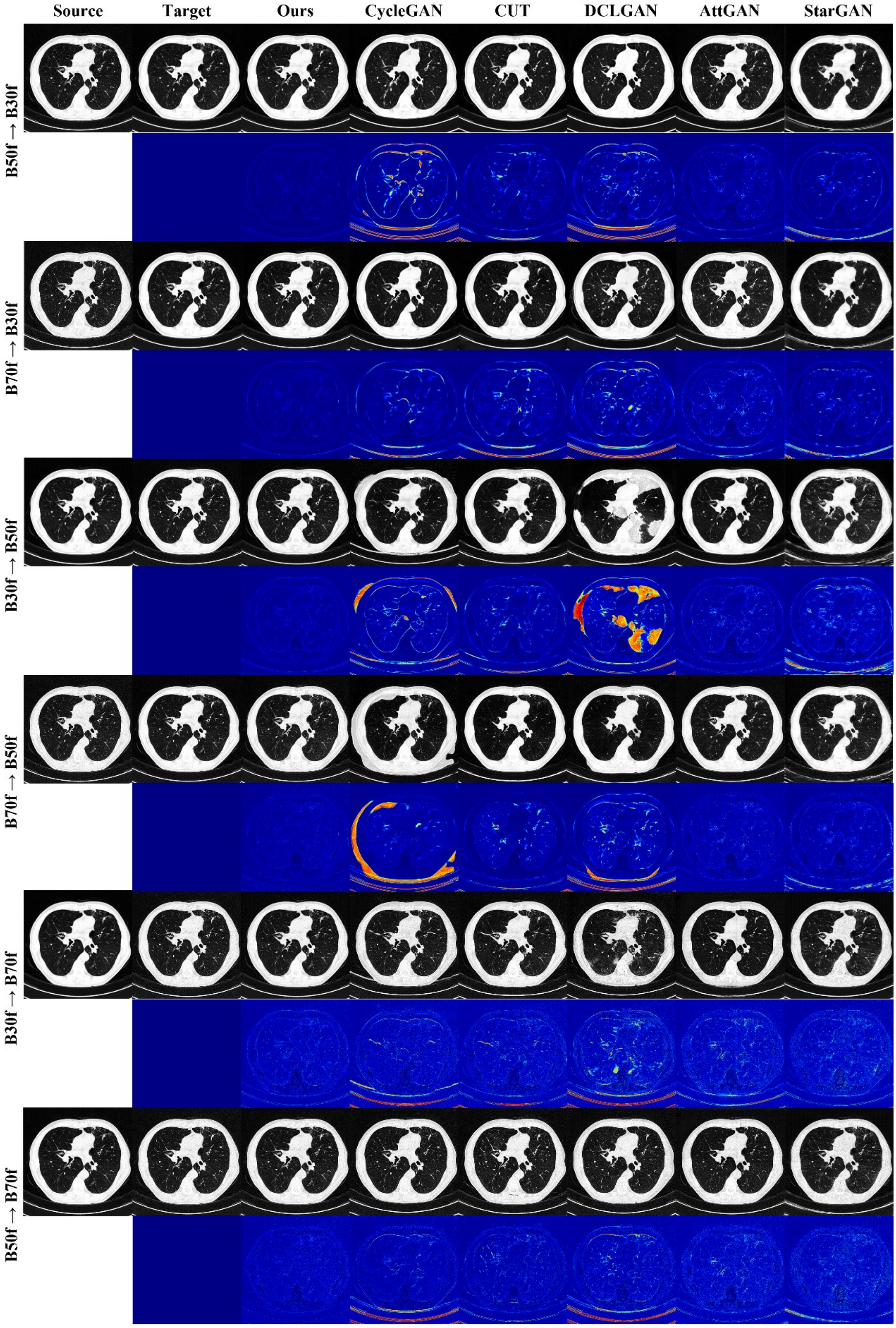

Figure 3. Visualization of pixel absolute value differences between the target images and the translated images. For each kernel conversion, the first row is the qualitative results of UNI2IT models including our proposed model (KCGAN). The second row is difference map corresponding to the first row.

As shown in Table 2, We also showed the quantitative results of kernel conversion about 6 directions. PSNR, SSIM and RMSE were used as quantitative metrics. In two domain UNI2IT models, CycleGAN and DCLGAN showed relatively low PSNR, SSIM and RMSE about all directions, but CycleGAN achieved 0.901 ± 0.021 in SSIM in the translation from B70f to B30f. CUT achieved 0.910 ± 0.024 as the best SSIM performance in the translation from B70f to B50f and comparably high PSNR, SSIM and RMSE about all directions. Nevertheless, it is still hard for them to translate the kernel from B30f to B70f. It should be noted that they are not robust and can't preserve the anatomical structures well. In multi-domain UNI2IT models, AttGAN showed great PSNR and RMSE performance, which is superior to StarGAN and CUT, excluding the translation from B30f to B70f. KCGAN showed tremendously increased PSNR, SSIM and RMSE performance compared to vanilla StarGAN and even other GAN-based UNI2IT models about all directions of kernel conversion. This indicates that our proposed method can help to improve robustness and performance when the model is learned about the kernel conversion.

Table 2. The quantitative results of kernel conversion about 6 directions.

Method	Sharp to Soft								
	B50f → B30f			B70f → B30f			B70f → B50f		
	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE
CycleGAN	26.522	0.833	195.512	29.662	0.901	139.161	24.882	0.790	238.286
	± 0.914	± 0.021	± 19.177	± 2.153	± 0.021	± 31.672	± 1.336	± 0.050	± 34.563
CUT	33.779	0.941	89.130	30.001	0.893	132.51	31.053	**0.910**	123.857
	± 2.585	± 0.020	± 27.837	± 1.816	± 0.029	± 25.402	± 2.150	**± 0.024**	± 28.234
DCLGAN	27.129	0.771	183.629	26.011	0.626	206.574	24.899	0.643	236.236
	± 1.491	± 0.023	± 30.120	± 0.801	± 0.046	± 18.485	± 0.972	± 0.044	± 24.910
AttGAN	39.211	0.941	45.213	35.832	0.889	66.639	34.761	0.817	75.627
	± 0.413	± 0.008	± 2.210	± 0.536	± 0.012	± 4.117	± 0.594	± 0.031	± 5.288
StarGAN	31.416	0.880	110.668	30.906	0.841	117.143	31.657	0.783	108.169
	± 0.859	± 0.028	± 10.672	± 0.636	± 0.040	± 8.377	± 0.982	± 0.035	± 11.791
KCGAN (ours)	**46.161**	**0.984**	**20.289**	**42.552**	**0.971**	**30.711**	**37.430**	0.897	**56.117**
	± 0.601	**± 0.006**	**± 1.488**	**± 0.592**	**± 0.010**	**± 2.167**	**± 0.707**	± 0.022	**± 4.745**

Method	Sharp to Soft								
	B30f → B50f			B30f → B70f			B50f → B70f		
	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE
CycleGAN	25.895	0.766	210.860	26.579	0.683	193.908	26.110	0.752	204.453
	± 1.104	± 0.054	± 25.015	± 0.952	± 0.039	± 21.624	± 0.623	± 0.021	± 14.743
CUT	31.700	0.867	108.520	28.884	0.694	149.090	30.762	0.796	120.877
	± 1.494	± 0.040	± 18.943	± 0.894	± 0.046	± 16.234	± 1.049	± 0.031	± 14.824
DCLGAN	25.557	0.719	218.819	25.209	0.604	237.335	25.830	0.652	214.752
	± 1.046	± 0.062	± 27.954	± 2.366	± 0.070	± 75.458	± 1.521	± 0.056	± 42.994
AttGAN	35.443	0.861	69.632	28.140	0.588	161.520	29.253	0.616	142.785
	± 0.384	± 0.020	± 3.121	± 0.550	± 0.039	± 10.199	± 0.663	± 0.040	± 11.235
StarGAN	28.774	0.755	149.692	28.456	0.604	156.015	27.355	0.584	177.083
	± 0.713	± 0.024	± 12.131	± 0.580	± 0.046	± 10.883	± 0.589	± 0.045	± 12.337
KCGAN (ours)	**38.728**	**0.925**	**47.948**	**31.735**	**0.764**	**107.787**	**32.721**	**0.797**	**97.135**
	± 0.571	**± 0.021**	**± 3.286**	**± 0.695**	**± 0.034**	**± 9.104**	**± 0.954**	**± 0.034**	**± 11.422**

Note—Mean and SD were calculated per patient; RMSE was calculated in range [-1024–3071].

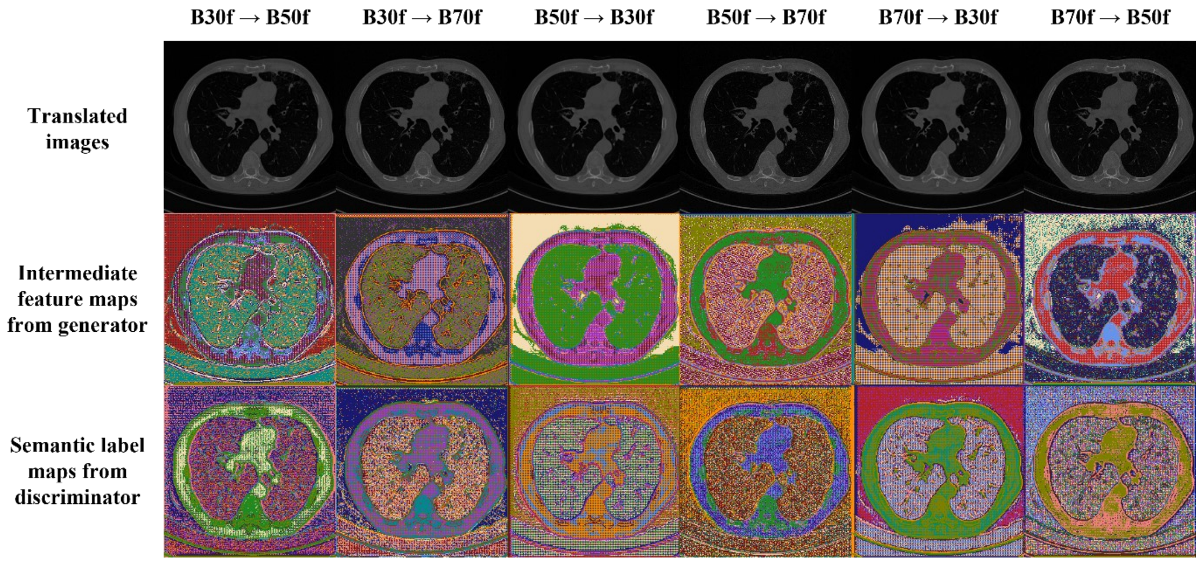

Figure 4. Visualization of the intermediate feature map from the generator and the semantic label map from the discriminator through k-means clustering ($k = 12$). The first row shows the translated CT image about 6 directions of kernel conversion. The second row shows the intermediate feature maps from the decoder of the generator. The third row shows the semantic label maps from the decoder of the discriminator. The size of the intermediate feature maps and the semantic label maps is 256 × 256.

Visualization of Intermediate Feature Map and Semantic Label Map

When using perceptual guidance, we can see how the intermediate feature map is extracted from the generator and the semantic label map is generated from the discriminator. For visualization of feature map, we clustered the pixel using k-means clustering. In this study, we used $k = 12$ for the coarse and fine-grained detail semantic information. As shown in Figure 4, the same anatomical structures such as bones with bones or muscles with muscles are clustered well in the intermediate feature map and the semantic label map. Interestingly, it was shown that the air in lung and the air outside are clustered differently. In addition, airway or vessels are also clustered differently with muscles which have similar pixel intensity. Through this, we could see that the generator could generate the translated image with rich semantic representation, so this intermediate feature map could teach the discriminator as a semantic ground-truth mask.

Ablation Studies

Ablation studies were implemented about the balance of the hyperparameters λ_{ccl} and λ_{csl}, separate usage of CCL and CSL, usage of encoder feature map from the generator for CCL, effect of perceptual guidance and computational cost difference between StarGAN and KCGAN. More details about the ablation studies are as follows:

1) *The balance of the hyperparameters* λ_{ccl} *and* λ_{csl}: we experimented the combination of the hyperparameters λ_{ccl} and λ_{csl} for the best performance. As shown in Table 3, when the combination

of CCL and CSL weights was 1 and 5, it showed the best performance, although 2 and 10 showed good performance in the translation from B70f to B30f.

2) *Separated usage of CCL and CSL*: we studied that whether only using one loss function—CCL or CSL—can improve performance. As shown in Table 4, the results showed that they can significantly improve performance. Especially when using CSL, it showed powerful performance in the translation from sharp kernels to soft kernels. Nevertheless, if we use perceptual guidance which uses CCL and CSL together, it is shown that the performance increased tremendously when translating from soft kernels to sharp kernels.

3) *Usage of encoder feature map from the generator for CCL*: cosine similarity content loss was defined to preserve coarse anatomical structure of the source image. CCL was utilized with the intermediate feature map from the decoder of the generator in our experiments, however, there is an encoder which can also extract the feature map. Thus, we experimented whether the intermediate feature map from encoder of the generator can help for the discriminator to learn semantic content information instead of the intermediate feature map from decoder. As shown in Table 5, although the encoder showed good performance, it was slightly lower than the decoder. This might be because the translated image has more similar content representation than the source image.

4) *Effect of perceptual guidance*: we experimented the effect of perceptual guidance in KCGAN. As shown in Table 6, it is shown that KCGAN without perceptual guidance and with cycle consistency loss already showed significantly increased performance compared to StarGAN. Nevertheless, KCGAN with perceptual guidance and cycle consistency loss showed that perceptual guidance can be effective in improving performance especially in translation from soft kernels to sharp kernels.

5) *Computational cost between baseline and our proposed method*: we calculated computational cost between baseline model (StarGAN) with and without architectural improvements and perceptual guidance to show efficiency of our proposed method. As shown in Table 7, the parameters and FLOPs increased 25.9% and 6.2%, respectively. This indicates that using our proposed method doesn’t cost a lot and is efficient. This experiment was conducted with image size 512 × 512 and a single NVIDIA GeForce RTX3090 24GB GPU.

Table 3. Ablation study about the balance of the hyperparameters λ_{ccl} and λ_{csl}.

Method	Sharp to Soft								
	B50f → B30f			B70f → B30f			B70f → B50f		
	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE
CSL 1 + CCL 5	**46.161** **± 0.601**	**0.984** **± 0.006**	**20.289** **± 1.488**	42.552 ± 0.592	0.971 ± 0.010	30.711 ± 2.167	**37.430** **± 0.707**	**0.897** **± 0.022**	**56.117** **± 4.745**
CSL 1 + CCL 10	45.313 ± 0.576	0.982 ± 0.006	22.332 ± 1.522	42.201 ± 0.624	0.970 ± 0.011	31.946 ± 2.364	36.198 ± 0.800	0.867 ± 0.031	59.652 ± 6.255
CSL 2 + CCL 5	44.353 ± 0.987	0.980 ± 0.004	25.095 ± 3.388	41.934 ± 0.715	0.967 ± 0.006	33.003 ± 3.029	36.182 ± 0.820	0.873 ± 0.024	64.472 ± 6.234
CSL 2 + CCL 10	45.359 ± 0.488	0.982 ± 0.005	22.263 ± 1.329	**43.213** **± 0.441**	**0.975** **± 0.007**	**28.447** **± 1.510**	36.349 ± 0.783	0.854 ± 0.030	63.411 ± 5.901
Method	Soft to Sharp								
	B30f → B50f			B30f → B70f			B50f → B70f		
	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE
CSL 1 + CCL 5	**38.728** **± 0.571**	**0.925** **± 0.021**	**47.948** **± 3.286**	**31.735** **± 0.695**	**0.764** **± 0.034**	**107.787** **± 9.104**	**32.721** **± 0.954**	**0.797** **± 0.034**	**97.135** **± 11.422**
CSL 1 + CCL 10	37.927 ± 0.562	0.909 ± 0.025	52.455 ± 3.526	31.401 ± 0.695	0.734 ± 0.036	111.941 ± 9.463	32.063 ± 0.899	0.760 ± 0.038	104.486 ± 11.512
CSL 2 + CCL 5	37.546 ± 0.472	0.912 ± 0.017	54.663 ± 3.113	31.414 ± 0.722	0.743 ± 0.037	111.921 ± 9.857	31.503 ± 0.935	0.737 ± 0.044	111.418 ± 12.782
CSL 2 + CCL 10	37.255 ± 0.635	0.889 ± 0.021	56.868 ± 4.318	30.619 ± 0.712	0.690 ± 0.039	122.826 ± 10.551	30.765 ± 0.878	0.692 ± 0.043	121.485 ± 12.941

Note—Mean and SD were calculated per patient; RMSE was calculated in range [-1024–3071]; CSL: Contrastive style loss; CCL: Cosine similarity content loss.

Table 4. Ablation study about separated usage of CCL and CSL.

Method	Sharp to Soft								
	B50f → B30f			B70f → B30f			B70f → B50f		
	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE
CCL	44.970 ± 0.482	0.978 ± 0.009	23.252 ± 1.363	42.655 ± 0.496	0.968 ± 0.014	30.325 ± 1.817	36.058 ± 0.831	0.847 ± 0.033	65.762 ± 6.518
CSL	45.668 ± 0.505	0.984 ± 0.007	21.466 ± 1.333	**43.398** **± 0.483**	**0.977** **± 0.009**	**27.830** **± 1.625**	**38.147** **± 0.662**	**0.898** **± 0.026**	**51.164** **± 4.064**
PG	**46.161** **± 0.601**	**0.984** **± 0.006**	**20.289** **± 1.488**	42.552 ± 0.592	0.971 ± 0.010	30.711 ± 2.167	37.430 ± 0.707	0.897 ± 0.022	56.117 ± 4.745
Method	Soft to Sharp								
	B30f → B50f			B30f → B70f			B50f → B70f		
	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE
CCL	37.704 ± 0.640	0.899 ± 0.021	54.012 ± 4.155	31.133 ± 0.789	0.718 ± 0.038	115.811 ± 11.007	31.235 ± 0.886	0.716 ± 0.039	114.926 ± 12.399
CSL	36.723 ± 0.392	0.893 ± 0.020	59.963 ± 2.772	30.961 ± 0.708	0.735 ± 0.033	117.735 ± 9.978	32.013 ± 0.908	0.775 ± 0.033	104.958 ± 11.643
PG	**38.728** **± 0.571**	**0.925** **± 0.021**	**47.948** **± 3.286**	**31.735** **± 0.695**	**0.764** **± 0.034**	**107.787** **± 9.104**	**32.721** **± 0.954**	**0.797** **± 0.034**	**97.135** **± 11.422**

Note—Mean and SD were calculated per patient; RMSE was calculated in range [-1024–3071]; CSL: Contrastive style loss; CCL: Cosine similarity content loss; PG: Perceptual guidance

Table 5. Ablation study about usage of encoder feature map from the generator for CCL.

Method	Sharp to Soft								
	B50f → B30f			B70f → B30f			B70f → B50f		
	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE
CCL from encoder	45.529 ± 0.506	**0.984** **± 0.005**	21.773 ± 1.335	42.073 ± 0.510	0.969 ± 0.011	32.410 ± 1.968	37.240 ± 0.911	0.884 ± 0.032	57.264 ± 6.324
CCL from decoder	**46.161** **± 0.601**	0.984 ± 0.006	**20.289** **± 1.488**	**42.552** **± 0.592**	**0.971** **± 0.010**	**30.711** **± 2.167**	**37.430** **± 0.707**	**0.897** **± 0.022**	**56.117** **± 4.745**
Method	Soft to Sharp								
	B30f → B50f			B30f → B70f			B50f → B70f		
	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE
CCL from encoder	38.450 ± 0.608	0.921 ± 0.022	49.426 ± 3.610	31.419 ± 0.687	0.743 ± 0.036	111.606 ± 9.266	32.222 ± 0.987	0.769 ± 0.041	102.578 ± 12.422
CCL from decoder	**38.728** **± 0.571**	**0.925** **± 0.021**	**47.948** **± 3.286**	**31.735** **± 0.695**	**0.764** **± 0.034**	**107.787** **± 9.104**	**32.721** **± 0.954**	**0.797** **± 0.034**	**97.135** **± 11.422**

Note—Mean and SD were calculated per patient; RMSE was calculated in range [-1024–3071]; CSL: Contrastive style loss; CCL: Cosine similarity content loss

Table 6. Ablation study about effect of perceptual guidance.

Method	Sharp to Soft								
	B50f → B30f			B70f → B30f			B70f → B50f		
	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE
StarGAN	31.416 ± 0.859	0.880 ± 0.028	110.668 ± 10.672	30.906 ± 0.636	0.841 ± 0.040	117.143 ± 8.377	31.657 ± 0.982	0.783 ± 0.035	108.169 ± 11.791
KCGAN	45.009 ± 0.604	0.969 ± 0.011	23.221 ± 1.677	**43.583** **± 0.547**	0.966 ± 0.010	**27.337** **± 1.792**	36.803 ± 0.836	0.867 ± 0.028	60.336 ± 6.066
w/ PG (ours)	**46.161** **± 0.601**	**0.984** **± 0.006**	**20.289** **± 1.488**	42.552 ± 0.592	**0.971** **± 0.010**	30.711 ± 2.167	**37.430** **± 0.707**	**0.897** **± 0.022**	**56.117** **± 4.745**

Method	Soft to Sharp								
	B30f → B50f			B30f → B70f			B50f → B70f		
	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE
StarGAN	28.774 ± 0.713	0.755 ± 0.024	149.692 ± 12.131	28.456 ± 0.580	0.604 ± 0.046	156.015 ± 10.883	27.355 ± 0.589	0.584 ± 0.045	177.083 ± 12.337
KCGAN	37.338 ± 0.662	0.892 ± 0.023	56.329 ± 4.496	30.539 ± 0.730	0.679 ± 0.041	124.017 ± 10.881	30.726 ± 0.831	0.683 ± 0.044	121.982 ± 12.312
w/ PG (ours)	**38.728** **± 0.571**	**0.925** **± 0.021**	**47.948** **± 3.286**	**31.735** **± 0.695**	**0.764** **± 0.034**	**107.787** **± 9.104**	**32.721** **± 0.954**	**0.797** **± 0.034**	**97.135** **± 11.422**

Note—Mean and SD were calculated per patient; RMSE was calculated in range [-1024–3071]; PG: Perceptual guidance.

Table 7. Ablation study about computational cost between baseline and our proposed method.

Method	#params	FLOPs
Baseline	54.1M	363.8G
+ Architecture improvements and perceptual guidance (ours)	68.1M	386.4G
	(+25.9%)	(+6.2%)

Discussion

In medical I2IT, preserving the anatomical structure of the source image is quite important as well as transferring the style of the target image because there are some cases that quantitative analysis is performed. Also, the changed anatomical structure is crucial when qualitative analysis is performed. However, GAN-based I2IT has a chronic problem that training procedure is not stable, so it makes to occur hallucination which generates fake structure. In UNI2IT, GAN-based networks rely on cycle consistency method generally. If this method fails to learn, the translated image can't preserve the structure like CycleGAN and StarGAN as shown in Figure 2. CUT, DCLGAN and AttGAN were proposed as new methods which can replace the cycle consistency method, but they also have the same limitation. We alleviated this problem with perceptual guidance by utilizing the intermediate feature map from the generator and the semantic label map from the discriminator and other regularization methods. In addition, through ablation study about effect of perceptual guidance, it is shown that using perceptual guidance and cycle consistency loss simultaneously can improve performance by complementing each other.

Perceptual guidance is motivated from GGDR [22] which uses cosine similarity loss between the intermediate feature map from the generator and the semantic label map from the discriminator. GGDR showed that it can improve fidelity as much as conditional GAN generation without any ground-truth semantic segmentation masks. However, the generator consists of an encoder and a decoder in I2IT differently from image generation and this is required that the generator should preserve the structure of the source input image. Especially for CT kernel conversion, preserving structure including fine-grained details while transferring style of the target image is important for qualitative and quantitative analysis, so only using cosine similarity loss may not be sufficient. Previous study [39] proposed PatchNCE loss instead of cosine similarity loss for the discriminator to learn fine-grained detail anatomical structure in CT kernel conversion. We thought that combination of cosine similarity loss and PatchNCE loss can help to learn overall coarse-to-fine anatomical structure including style of the target image in feature domain.

The advantage of perceptual guidance was shown in the translation from soft kernels to sharp kernels. Despite it is a more difficult task because the spatial resolution should be increased, and the noise pattern should be clear, perceptual guidance showed great performance. Also, perceptual guidance showed tremendous improvement through quantitative results in 6 directions of the kernel conversion. Furthermore, to apply perceptual guidance, it requires no additional learnable or pre-trained encoding networks and low computational cost. Nevertheless, our study has some limitations. First, KCGAN showed weakness in qualitative results. It could preserve the coarse-to-fine details anatomical structure, but it still could not transfer the style of the target image perfectly as shown in Figure 2. Second, we did not show the results about the translation of more various kernels including external manufacturers, so our proposed method needs to prove generalizability of kernel conversion. Finally, we didn't compare our proposed method with denoising diffusion probabilistic models (DDPM) [40] which show tremendous generation performance better than GAN, recently.

Conclusion

In this study, we proposed perceptual guidance which regularizes the discriminator for robust and efficient learning of GAN-based multi-domain image-to-image translation. Our proposed method can preserve the coarse-to-fine detail anatomical structure of the source image. This method needs only changing discriminator architecture to U-Net and does not require introducing any additional learnable or pre-trained networks which are accompanied by an encoding process. Experimental results showed that our proposed method outperformed existing GAN-based image-to-image translation models in CT kernel conversion.

References

1. Mackin, D., et al.: Matching and homogenizing convolution kernels for quantitative studies in computed tomography. Invest. Radiol. **54**(5), 288 (2019).

2. Lee, S.M., et al.: CT image conversion among different reconstruction kernels without a sinogram by using a convolutional neural network. Korean J. Radiol. **20**(2), 295–303 (2019).

3. Choe, Jooae, et al.: Deep learning–based image conversion of CT reconstruction kernels improves radiomics reproducibility for pulmonary nodules or masses. Radiology **292**(2), 365-373 (2019).

4. Pang, Y., et al.: Image-to-image translation: methods and applications. IEEE Trans. Multimedia **24**, 3859–3881 (2021).

5. Eun, D.-I., et al.: CT kernel conversions using convolutional neural net for super-resolution with simplified squeeze-and-excitation blocks and progressive learning among smooth and sharp kernels. Comput. Meth. Programs Biomed. **196**, 105615 (2020).

6. Bak, So Hyeon, et al.: Emphysema quantification using low-dose computed tomography with deep learning–based kernel conversion comparison. European Radiology **30**, 6779-6787 (2020).

7. Goodfellow, I., et al.: Generative adversarial networks. Commun. ACM **63**(11), 139–144 (2020).

8. Gravina, M., et al.: Leveraging CycleGAN in Lung CT Sinogram-free Kernel Conversion. In: Sclaroff, S., Distante, C., Leo, M., Farinella, G.M., Tombari, F. (eds.) Image Analysis and Processing – ICIAP 2022: 21st International Conference, Lecce, Italy, May 23–27, 2022, Proceedings, Part I, pp. 100–110. Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-031-06427-2_9.

9. Yang, S., Kim, E.Y., Ye, J.C.: Continuous conversion of CT kernel using switchable CycleGAN with AdaIN. IEEE Trans. Med. Imaging **40**(11), 3015–3029 (2021).

10. Kong, L., et al.: Breaking the dilemma of medical image-to-image translation. Adv. Neural. Inf. Process. Syst. **34**, 1964–1978 (2021).

11. Isola, P., et al.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017).

12. Zhu, J.-Y., et al.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision (2017).

13. Choi, Y., et al.: Stargan: Unified generative adversarial networks for multi-domain image-toimage translation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2018).

14. He, Z., et al.: Attgan: facial attribute editing by only changing what you want. IEEE Trans.

Image Process. **28**(11), 5464–5478 (2019).

15. Liu, M.-Y., Breuel, T., Kautz, J.: Unsupervised image-to-image translation networks. Adv. Neural Inform. Process. Syst. **30** (2017).

16. Park, T., et al. (eds.): Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX, pp. 319–345. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-58545-7_19.

17. Han, J., et al.: Dual contrastive learning for unsupervised image-to-image translation. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 746–755 (2021).

18. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE International Conference on Computer Vision (2017).

19. Shi, Wenzhe., et al.: Differentiable augmentation for data efficient gan training. Advances in Neural Information Processing Systems 33, 7559–7570 (2020).

20. Gulrajani, I., et al.: Improved training of wasserstein gans. Adv. Neural Inform. Process. Syst. **30** (2017).

21. Miyato, T., et al.: Spectral normalization for generative adversarial networks. arXiv preprint arXiv:1802.05957 (2018).

22. Lee, G., et al.: Generator knows what discriminator should learn in unconditional GANs. In: Avidan, Shai, Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVII, pp. 406–422. Springer Nature Switzerland, Cham (2022). https://doi.org/10.1007/978-3-031-19790-1_25.

23. Arjovsky, M., et al.: Wasserstein generative adversarial networks. In ICML, 214–223, (2017).

24.loffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. International conference on machine learning. pmlr, 448–456, (2015).

25. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411. 1784 (2014).

26. Park, T., et al.: Semantic image synthesis with spatially-adaptive normalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2019).

27. Sushko, V., et al.: You only need adversarial supervision for semantic image synthesis. arXiv preprint arXiv:2012.04781 (2020).

28. Wang, X., et al.: Dense contrastive learning for self-supervised visual pre-training. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2021).

29. Xie, Z., et al.: Propagate yourself: exploring pixel-level consistency for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2021).

30. He, K., et al.: Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. (2016).

31. Gatys, L. A., et al.: Image style transfer using convolutional neural networks. Proceedings of the IEEE conference on computer vision and pattern recognition. (2016).

32. Johnson, J., et al.: Perceptual losses for real-time style transfer and super-resolution. Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14. Springer International Publishing. (2016).

33. Simonyan, K., et al.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014).

34. Zhang, Y., Yang, Q.: A survey on multi-task learning. IEEE Trans. Knowl. Data Eng. **34**(12), 5586–5609 (2021).

35. Kingma, D.P, Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412. 6980 (2014).

36. Shi,W., et al.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016).

37. Odena, A., Dumoulin, V., Olah, C.: Deconvolution and checkerboard artifacts. Distill. **1**(10), e3 (2016).

38. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18. Springer (2015).

39. Choi, C., et al.: CT Kernel Conversion Using Multi-domain Image-to-Image Translation with Generator-Guided Contrastive Learning. International Conference on Medical Image Computing and Computer Assisted Intervention. Cham: Springer Nature Switzerland. 344-354, (2023).

40. Ho, Jonathan., et al.: Denoising diffusion probabilistic models. Advances in neural information processing systems. 33, 6840–6851 (2020).

Abstract (with Korean)

컴퓨터 단층촬영 (CT) 영상은 의료 분야에서 널리 사용되는 진단 영상 중 하나이다. CT 영상은 투영을 포함하는 2차원 배열 데이터인 사이노그램으로부터 역투영을 통한 컨볼루션 커널을 이용하여 재구성된다. 이때 정성적 평가에서는 어떤 해부학적 구조는 평가하느냐에 따라 커널이 달라진다. 또한 정성적 평가뿐만 아니라 정량적 평가도 중요하며 커널 선택에 영향을 미친다. 그러나 여기에는 두가지 문제가 있다. 첫번째로, 사이노그램은 용량이 크고 저장 공간은 제한되어 있기 때문에 일반적으로 평가를 위해 하나의 특정 커널만으로 CT 영상을 재구성하고 일주일 내에 사이노그램을 제거한다. 두번째로, 환자를 스캔하고 다시 한번 방사선에 노출이 된다. 최근에는 많은 연구자들이 CT 커널 변환을 위해 적대적 생성 신경망 (GAN)을 사용한 이미지 대 이미지 변환 방법을 제안해왔다. 그럼에도 불구하고, CT 이미지가 소스 커널에서 대상 커널로 변환될 때 기도 및 혈관과 같은 미세한 세부 사항을 포함한 해부학적 구조를 보존하면서 대상 커널의 스타일로 변환하는 것은 여전히 어려운 일이다. 본 연구에서는 이러한 문제를 지각적 안내로 완화하기 위해 커널 변환 GAN (KCGAN)을 제안하고 커널 변환에서 강력하고 효율적인 성능을 보여주었다. 지각적 안내는 의미론적 표현을 더 잘 학습하기 위해 생성자의 특징 맵을 사용하는 판별자 정규화 방법의 일종이다. 콘텐츠 및 스타일 특징의 경우 생성자의 특징 맵과 판별자의 의미 레이블 맵 간에 코사인 유사성 콘텐츠 손실과 대비 스타일 손실이 각각 정의된다. KCGAN은 소스 도메인의 세밀한 해부학적 구조를 보존하는 동시에 대상 도메인의 스타일을 전달할 수 있다. 또한 이 방법은 판별자 구조만 변경하고 추가적인 학습 가능 네트워크나 사전 훈련된 네트워크를 활용하지 않고도 쉽게 적용할 수 있다. 실험 결과, 3개 커널 사이에서 대부분의 커널 변환 방향에서 이 방법이 기존 GAN 기반 방법보다 성능이 뛰어난 것을 보여주었다.