Neural Contrast Enhancement of CT Image

Minkyo Seo[1] Dongkeun Kim[1] Kyungmoon Lee[1] Seunghoon Hong[2]
Jae Seok Bae[3] Jung Hoon Kim[4] Suha Kwak[1]

POSTECH[1] KAIST[2] Seoul National University Hospital[3]
Seoul National University College of Medicine[4]

Abstract

Contrast materials are often injected into body to contrast specific tissues in Computed Tomography (CT) images. Contrast Enhanced CT (CECT) images obtained in this way are more useful than Non-Enhanced CT (NECT) images for medical diagnosis, but not available for everyone due to side effects of the contrast materials. Motivated by this, we develop a neural network that takes NECT images and generates their CECT counterparts. Learning such a network is extremely challenging since NECT and CECT images for training are not aligned even at the same location of the same patient due to movements of internal organs. We propose a two-stage framework to address this issue. The first stage trains an auxiliary network that removes the effect of contrast enhancement in CECT images to synthesize their NECT counterparts well-aligned with them. In the second stage, the target model is trained to predict the real CECT images given a synthetic NECT image as input. Experimental results and analysis by physicians on abdomen CT images suggest that our method outperforms existing models for neural image synthesis.

1. Introduction

As a medical imaging tool, Computed Tomography (CT) has been employed to take a sequence of cross-sectional images of human body for a wide range of clinical purposes. When taking CT scans, *contrast materials* are often injected into body to improve the visibility of specific organs, blood vessels, or tissues by enhancing contrast between such areas and surrounding structures in CT images. This approach is called Contrast Enhanced CT (CECT), and presents useful anatomical information that cannot be captured by the ordinary Non-Enhanced CT (NECT) administering no contrast material. Compared to NECT, however, CECT is costly, demands more radiation exposure, and may cause side effects such as vomiting and headache. Furthermore, CECT would be risky for patients with kidney diseases or having allergies to contrast materials.

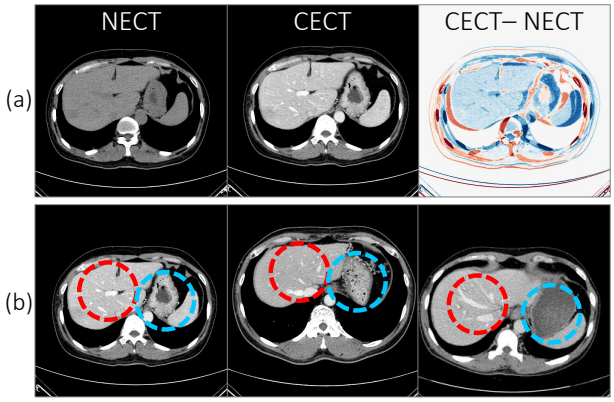

Figure 1. Two main challenges in our task. (a) NECT and CECT images taken at the same location of the same patient can be misaligned. Thus, a large portion of intensity changes between the images is caused by the misalignment and irrelevant to the effect of contrast materials. (b) Contrast enhancement patterns in CECT images are not consistent across patients, but vary significantly according to their medical conditions.

Motivated by this, we aim to develop a framework that helps physicians better diagnose medical conditions in abdomen CT images without the disadvantages of CECT. We study a data-driven approach that synthesizes a CECT image corresponding to the given NECT image without introducing contrast materials. To this end, we first collect a set of NECT and CECT images taken before and after injecting contrast materials, respectively, then train deep neural networks to learn the mapping from NECT to CECT images of the same patient in the collected dataset.

Our target task is thus a neural image synthesis problem, but there are two main challenges that differentiate the task from existing problems like style transfer [8, 9, 10, 18, 22, 24, 25, 36] and image-to-image translation [16, 19, 27, 30, 34, 44]. First, NECT and CECT images of the same patient are often largely misaligned as shown in Fig. 1(a) due to morphological distortions caused by peri-

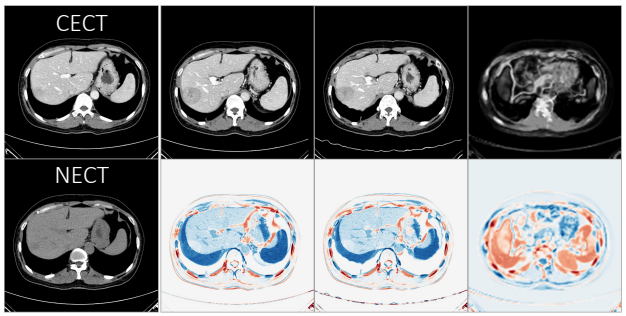

Figure 2. Failure examples of registration. (a) Affine transform. (b) B-spline. (c) VoxelMorph [2], a voxel registration technique based on deep unsupervised learning. Top images are registration results of the methods and bottom images visualize differences between the registered images and the real NECT images.

staltic and respiratory movements. We empirically found that this misalignment issue is hard to be addressed by conventional image registration techniques [2, 17, 21, 28] as demonstrated in Fig. 2 due to the severe intensity variations and complicated distortions between NECT and CECT images. Hence a direct supervision for the mapping from NECT to CECT is not accessible. Second, aspects of contrast enhancement in CECT images vary greatly across patients with different medical conditions as illustrated in Fig. 1(b). It is thus not straightforward to define a common style of the CECT domain, and existing neural style transfer methods could have trouble recognizing and contrasting specific areas affected by contrast materials since they are designed to transfer *domain-specific* styles rather than *example-specific* (*patient-specific* in our task) information in general.

To address the above challenges, we propose a two-stage framework. In the first stage, we train an auxiliary network that takes real CECT images and synthesizes their NECT counterparts by removing the effect of contrast enhancement in the input CECT images; a pair of real CECT and synthetic NECT images obtained in this stage are aligned almost perfectly. We argue that the first stage, an inverse of our target task, is more feasible than the target task since NECT images are much less patient-specific than CECT images due to their monotonic intensities on the areas of internal organs. It is thus easier to learn a common style of the NECT domain and transfer the style to CECT images without aligned NECT-CECT pairs. Then in the second stage, our target model is trained to predict the real CECT images when the corresponding synthetic NECT images are given as input. Hence the target model trained in this stage can enjoy the strong patient-specific supervision based on reconstruction losses thanks to the aligned pairs of synthetic NECT and real CECT images.

The efficacy of our framework is evaluated on real abdomen CT images. According to evaluations by physicians, our method is better than existing models for neural image synthesis in terms of its high image quality and low degree of artifact. Due to the misalignment issue, unfortunately, it is not straightforward to evaluate quantitative performance precisely on the CT images. For this reason, we employ the IXI brain MRI dataset for additional experiments, in which our method and the existing models learn the mapping between two different domains of brain images in the presence of simulated distortions between the domains. Our method outperforms the baseline models quantitatively in these experiments. The contribution of this paper is three-folds.

- We introduce a new and challenging medical image synthesis task to the computer vision community.
- We propose the two-stage framework that is carefully designed to address the main challenges in the task.
- Physicians reported that our method is more useful than existing image synthesis models in clinical use.

2. Related Work

2.1. Image-to-Image Translation

Image-to-image translation aims to convert an image in one domain to another, such as sketch to photo [16, 43], label to pixel [38], masked to complete image [14]. Recently, conditional Generative Adversarial Network (cGAN) [5, 16, 38] have shown to be effective in this task. Isola *et al.* [16] employ a convolutional encoder-decoder network with an adversarial loss to learn a mapping between paired images. To ensure the alignment between input and output, it also adopt a regression loss based on L_1 distance between groundtruth and the predicted image. Later methods improve the quality of generated images by employing a stronger regression loss, such as the perceptual loss using the pre-trained classifier [5] or the feature-matching loss using multi-scale discriminators [38]. However, training these models requires many pairs of input and output data, which are not often available, especially in medical images.

To alleviate this limitation, unpaired image-to-image translation techniques have been proposed [11, 19, 23, 44]. In particular, CycleGAN [44] achieves the goal by encouraging a generator to create an output that can be inverted back into the input image by another generator. These approaches have demonstrated great success in many applications of image-to-image translation, but often generate artifacts looking plausible yet incorrect. In the medical image domain, such artifacts could be fatal since they may disturb correct diagnosis or distort inherent properties of a subject [6]. Qualitative examples of such artifacts generated by CycleGAN can be found in Fig. 6.

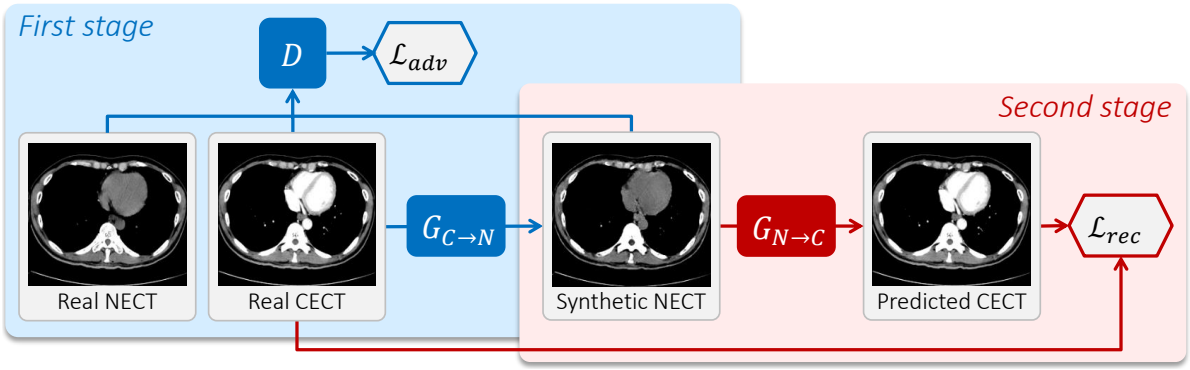

Figure 3. Overview of our two-stage framework. The first stage, colored in blue, trains the auxiliary network $G_{C\to N}$ to synthesize a realistic NECT image corresponding to the input CECT image. In the second stage colored in red, our target model denoted by $G_{N\to C}$ is trained to predict the real CECT image paired with the input synthetic NECT through reconstruction losses.

2.2. Neural Style Transfer

Our task shares the similar objective with the task of neural style transfer, which aims to transfer the style of one image to another while preserving its content [8, 9, 18, 22]. Existing methods manipulate the target image to match its feature statistics to that of reference image by iterative optimization [8, 9] or a learned feed-forward network [18, 22].

These methods unfortunately cannot be directly applied to our problem due to the absence of reference CECT images. In the case of iterative optimization [8, 9], reference CECT images of a patient are not available as they are the unknown targets we aim to predict. Also, it is impractical to utilize the feed-forward network [18, 22] since the style of CECT images vary significantly between different patients as can be seen in Fig. 1; due to the style gap between the reference CECT for training and latent target CECT in testing, this method will be likely to augment an inappropriate style to the input image, which could be fatal in our task since an incorrectly translated image misleads physicians and prevents precise diagnosis consequently.

2.3. Medical Image Synthesis

Medical image synthesis is an active research topic with many useful applications such as medical image denoising [4], data augmentation [12, 13], and cross-modality image synthesis [29, 37, 41]. Recently proposed cross-modality medical image synthesis methods [7, 33, 40, 41] are based on the conditional GANs [27] or CycleGAN [44]. In particular, MRI-to-CT techniques [29, 39] generate a CT image corresponding to the given MRI image so as to obtain CT images without the risk of radiation exposure. However, these techniques usually work on brain images, which are well aligned between different modalities unlike abdomen CT images. For this reason, our problem requires a method that is more robust to misalignment between source and target images than the image-to-image translation methods.

3. Our Approach

There are two main challenges in our task. First, NECT and CECT images are not aligned even at the same location of the same patient. Second, it is not straightforward to learn a common style of CECT images since aspects of contrast enhancement in CECT images vary greatly across patients. It is thus not straightforward to learn the mapping from NECT to CECT directly from real CT images.

We propose a two-stage framework to address these issues. The key idea is to *synthesize* well-aligned NECT and CECT image pairs to train our target model using conventional reconstruction losses. To this end, in the first stage we train an auxiliary network that removes the effect of contrast materials in a real CECT image, and utilize the network to generate pairs of synthetic NECT and real CECT images aligned to each other. In other words, the first stage learns the inverse of our task, which is more feasible to achieve than the target task due to the monotonic and less patient-specific appearances of NECT images. In the second stage, the pairs of aligned CT images are used to train our target model that predicts a CECT image corresponding to the input NECT image. Since the input and groundtruth images are aligned in this stage, our target model can enjoy the patient-specific supervision by reconstruction losses.

An overview of our approach is presented in Fig. 3, and the remaining part of this section discusses each of the two stages of our framework, network architectures, the design choice in more details.

3.1. First Stage

In this stage, we train an auxiliary network, denoted by $G_{C\to N}$, that takes a real CECT image and generates a synthetic NECT image corresponding to the input. The network is trained jointly with a discriminator D in an adversarial manner. A common choice of D is a binary classifier that discriminates real NECT and synthetic NECT images,

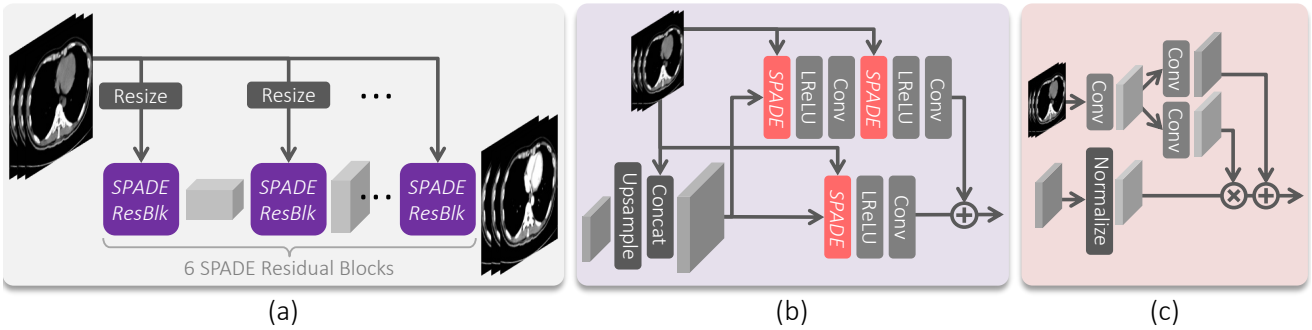

Figure 4. Illustration of our generators. (a) Overall architecture of our generators. (b) SPADE residual block. (c) SPADE module.

but motivated by AM-GAN [42], our D takes and classifies real CECT images as well so that it more actively enforces $G_{C\to N}$ to remove contrast enhancement patterns in the input CECT image. We thus employ as the discriminator D a three-class classifier that discriminates real NECT, real CECT, and synthetic NECT images at the same time.

$G_{C\to N}$ and D are then trained by optimizing the following two objectives alternately:

$$\min_{D} \; E_{x\sim p_C(x)}\big[H(\mathbb{1}_C, D(x)) + H(\mathbb{1}_S, D(G_{C\to N}(x)))\big] + E_{x\sim p_N(x)}\big[H(\mathbb{1}_N, D(x))\big], \tag{1}$$

$$\min_{G_{C\to N}} \; E_{x\sim p_C(x)}\big[H(\mathbb{1}_N, D(G_{C\to N}(x)))\big], \tag{2}$$

where C, N, and S indicate real CECT, real NECT, and synthetic NECT classes, respectively. Also, H is cross-entropy and $\mathbb{1}_k \in \mathbb{R}^3$ is a one-hot vector of the class $k \in \{C, N, S\}$. By learning $G_{C\to N}$ and D jointly in this manner, $G_{C\to N}$ becomes capable of generating synthetic NECT images that look realistic and have no contrast enhancement pattern at the same time.

3.2. Second Stage

In the second stage, our target model for neural contrast enhancement, denoted by $G_{N\to C}$, is learned to predict the real CECT image when the corresponding synthetic NECT image is given as input. Since the synthetic NECT and real CECT images are well-aligned, we can employ reconstruction losses to train $G_{N\to C}$. To this end, we first adopt $L1$ loss, which is defined as:

$$\mathcal{L}_1 = E_{x\sim p_C(x)}\big[||G_{N\to C}(G_{C\to N}(x)) - x||_1\big]. \tag{3}$$

Since the $L1$ loss often produces results perceptually unsatisfactory in terms of image quality, we also employ the perceptual loss [18]:

$$\mathcal{L}_{pcp} = E_{x\sim p_C(x)}\left[\sum_i w_i \frac{||\phi_i(x) - \phi_i(G_{N\to C}(G_{C\to N}(x)))||_1}{W_i H_i C_i}\right], \tag{4}$$

where ϕ_i denotes the feature map extracted from the i^{th} intermediate layer of a pretrained network, w_i indicates a balancing coefficient for ϕ_i, and W_i, H_i, and C_i indicate the width, height, and the number of channels of the feature map, respectively. As the pretrained network computing ϕ_i, we adopt a VGG16 network [35] with batch normalization [15] that is trained for classifying NECT and CECT images in our training dataset. Specifically, we utilize feature from the first four max-pooling layers. Finally, $G_{N\to C}$ is trained by minimizing the following objective:

$$\mathcal{L}_{rec} = \lambda\mathcal{L}_1 + \mathcal{L}_{pcp}, \tag{5}$$

where λ is a balancing coefficient.

3.3. Architectures of G and D

For both of the two generators $G_{C\to N}$ and $G_{N\to C}$ in our framework, we adopt the architecture of SPADE [30], one of the state-of-the-art in image-to-image translation. Specifically, the generators are built by stacking six SPADE residual blocks, each of which is followed by a bilinear upsampling operation. Also, they take as input and produce as output three consecutive CT images at once to capture 3-dimensional contexts. The input CT images are fed to each SPADE residual block, and also concatenated to the output of each upsampling operation. The overall architecture of our generators is illustrated in Fig. 4.

Meanwhile, for D of the first stage, we adopt the discriminator of DCGAN [32] and replace its trainable downsampling layers with bilinear interpolations.

3.4. Discussion

Advantage of the two-stage framework. The key advantage of our framework is that it can provide the pixel-level patient-specific supervision to our target model in the second stage. Note that the same model learned directly with real NECT-CECT image pairs by an adversarial loss often produces artifacts not present in the real CECT image as shown in Fig. 6 (*Single*). Such artifacts are fatal in our task since they may lead to wrong medical diagnosis and treatment. A main source of this problem is the adversarial

loss that provides a domain-level supervision only, which enables to learn typical contrast enhancement patterns in CECT images without aligning NECT and CECT images, but cannot inform which areas should be contrasted in a specific NECT image. On the other hand, the reconstruction losses in the second stage provide a patient-specific supervision in a pixel-level, which allows $G_{N\to C}$ to localize and contrast specific areas affected by contrast materials and alleviates the artifact problem in consequence.

Why $G_{C\to N}$ can be trained adversarially while $G_{N\to C}$ cannot. Since NECT images have monotonous intensities and textures in common, their styles are consistent and can be easily captured by $G_{C\to N}$ learned in an adversarial manner. On the other hand, contrast enhancement patterns in CECT images vary significantly across patients and cannot be accurately modeled by the weak domain-level supervision that the adversarial loss provides.

Why synthetic NECT and real CECT are aligned. There are two reasons why a synthetic NECT image predicted by $G_{C\to N}$ is well-aligned to the input CECT image. First, the generator does not need to deform the input image to cheat D since the goal can be achieved simply by reducing intensities of a few small areas affected by contrast materials. Second, the SPADE architecture of $G_{C\to N}$ inherently prevents distortion of the input image since the image is fed to every SPADE residual block during the generation procedure and parameters of $G_{C\to N}$ are trained for only small modifications accordingly.

Why not using the discriminator of SPADE. We use the discriminator of DCGAN, instead of that of SPADE, since the SPADE discriminator is not suited to our task. This model is designed to classify pairwise relations of input and output images into two categories, thus in our case, it discriminates between (real CECT, real NECT) and (real CECT, synthetic NECT). Since every pair of real NECT and CECT images undergoes morphological distortions, the discriminator considers such distortions as a property of real NECT images and forces $G_{C\to N}$ to synthesize distorted NECT images, which cannot be used as input to the second stage.

4. Experiments

The effectiveness of the proposed framework is evaluated and compared with existing models for neural image synthesis on the abdomen CT image dataset we collected. In addition, we employ the IXI brain MRI dataset for further performance analysis since it is tricky to precisely quantify the accuracy of the models in the CT image dataset due to the misalignment between NECT and CECT images.

The rest of this section first describes details of implementation, baseline methods, and evaluation metrics, then presents experimental results on the abdomen CT image dataset and the IXI brain MRI dataset.

4.1. Implementation Details

In our generators, outputs of the SPADE residual blocks have 512, 512, 256, 128, 64, and 32 channels. The convolution layers of the discriminator consist of 64, 128, 256, 512, 1024, 2048 channels. For both of the generators and the discriminator, we apply group normalization to all convolution layers and adopt leaky ReLU [26] with a negative slope of 0.2. The coefficients w_i in Eq. (4) are set to $\frac{1}{32}, \frac{1}{16}, \frac{1}{8}, \frac{1}{4}$ for $i = 1, 2, 3, 4$, respectively. Also, λ in Eq. (5) is set to 10. Our models were implemented in PyTorch [31], and optimized by ADAM [20] with $\beta_1 = 0.5$, $\beta_2 = 0.999$, and mini-batches of 4 images for 30 epochs on the abdomen CT image dataset and for 20 epochs on the IXI brain MRI dataset. The learning rate was initially 0.0001 and decayed by 0.9 at every epoch.

4.2. Baselines and Two Versions of Our Method

The proposed method is compared with two baselines: a single stage framework (Single) and CycleGAN (Cycle). Single trains the target model $G_{N\to C}$ jointly with the discriminator D through an adversarial loss only. Meanwhile, Cycle directly follows the original CycleGAN [44] training $G_{N\to C}$ and $G_{C\to N}$ jointly with discriminators. The difference between Cycle and the original one is that Cycle employs both of $L1$ and perceptual losses for the cycle consistency while the original one utilizes $L1$ loss only. For a fair comparison, both baselines are implemented by the same network architectures introduced in Sec. 3.3.

Since Single trains $G_{N\to C}$ only with the domain-level supervision, it produces artifacts frequently and cannot capture patient-specific patterns of contrast enhancement accurately. Likewise, in Cycle, $G_{N\to C}$ is trained by an adversarial loss as well as the cycle consistency loss, thus is prone to produce artifacts as in Single. Further, $G_{C\to N}$ of Cycle is trained with synthetic CECT images as well as the real ones, thus the quality of synthetic NECT images generated by the model prone to be more degraded compared to $G_{C\to N}$ of our model trained with real CECT images only.

In addition, we design two different versions of our method with two distinct training strategies: ours trained jointly (Ours-J) and ours trained separately (Ours-S). Ours-J trains both of $G_{C\to N}$ and $G_{N\to C}$ of our two-stage framework jointly in an end-to-end manner. On the other hand, Ours-S learns the two generators one by one, *i.e.*, it first optimizes $G_{C\to N}$ then trains $G_{N\to C}$ with the frozen $G_{C\to N}$. Ours-J is a natural training strategy, while Ours-S allows $G_{C\to N}$ to focus soley on generating realistic NECT images without being distracted by $G_{N\to C}$. We believe that Ours-S reduces the domain gap between real and synthetic NECT images, and could improve the performance of the target model $G_{N\to C}$ in consequence. The difference between Ours-S and Ours-J in accuracy is marginal as summarized in Tab. 1, but the results of Ours-S were in general

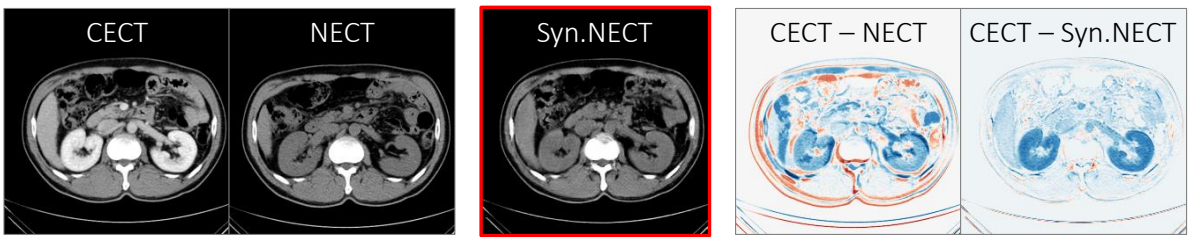

Figure 5. A synthetic NECT image (Syn.Nect) and its comparison to real NECT image.

	Abdomen CT Image		IXI brain MRI					
			$k = 0.01$		$k = 0.02$		$k = 0.05$	
	MSE	MS-SSIM	MSE	MS-SSIM	MSE	MS-SSIM	MSE	MS-SSIM
Single	0.108	0.533	0.356	0.928	0.484	0.914	0.574	0.907
Cycle	0.112	0.528	0.317	0.937	0.351	0.929	**0.318**	0.930
Ours-J	**0.097**	0.557	0.298	**0.950**	**0.312**	0.945	0.367	**0.941**
Ours-S	0.099	**0.559**	**0.271**	0.949	0.313	**0.946**	0.364	0.936

Table 1. Quantitative results on the IXI brain MRI dataset and the abdomen CT image dataset. k indicates the degree of distortion applied to training images; larger k means larger distortion (see Sec. 4.4 for details). We scale up MSE on the IXI brain MRI dataset 100 times to show performance gaps more clearly.

better than those of Ours-J in the perceptual quality on the abdomen CT image dataset as can be seen in Fig. 6 and 7.

4.3. Results on Abdomen CT Image Dataset

4.3.1 Dataset Specification

NECT and CECT abdomen images of our dataset are taken from 327 patients before and after injecting contrast materials, respectively. In consequence, we collect in total 23,923 pairs of abdomen NECT and CECT images. Among them, 19,180 pairs of 262 patients are used for training and remaining 4,743 pairs of 65 patients are kept for quantitative evaluation. In addition, we prepare extra 1,920 NECT images of other 16 patients for evaluation by physicians in terms of the image quality and the degree of artifact.)

All the images are of 256×256 resolution, where 1 millimeter in the real world corresponds to roughly 0.75 pixel.

The CT images are then converted into grayscale images for the convenience of processing and visualization. To this end, we adopt the windowing technique [3] and follow the common practice in this setting: 300 HU for window width and 50 HU for window level, where HU denotes Hounsfield units. Specifically, the interval of CT pixel values $[-100, 200]$ in HU is linearly transformed to that of grayscale intensities $[0, 255]$, and CT pixel values outside of the interval are clamped to 0 or 255 after the transformation.

4.3.2 Performance Analysis

We first present examples of synthetic NECT image and compare it with real NECT image in Fig. 5. It can be seen from the last difference images that false contrast changes (red) are reduced significantly while correct contrast enhancement patterns (blue) are correctly captured in our synthetic NECT images, which enable us to train the target model with the pixel-level reconstruction losses.

Qualitative results of our final models and the baselines on the abdomen CT image dataset are presented in Fig. 6. Single and Cycle, which rely on an adversarial loss for learning $G_{N\to C}$, produce noticeable artifacts frequently as shown in Fig. 6(a-c). On the other hand, our models rarely generate artifacts yet produce slightly blurry images in general. The results of Ours-S and Ours-J look similar, but Ours-S is slightly better than Ours-J in the quality of fine-grained details, *e.g.*, clearer enhancement in Fig. 6(a-c). Further, all methods successfully contrast organs and vessels located at regular positions, but for small lesions or tissues whose locations vary across the patients, the baselines could not capture the subtle patterns as much as our models could. In addition, all methods failed to enhance contrast for blood vessels in liver as shown in Fig. 6(d). It is highly challenging to recognize the vessels since their structures are substantially diverse. Quantitative results of the models are summarized in Tab. 1, where our models outperform the baselines in all metrics.

We further evaluate the effectiveness of our method in clinical use. To this end, we prepare extra 1,920 NECT images of 16 patients, without their CECT counterparts. Then physicians evaluate the quality of predicted CECT images and assign one of grades from 1 (excellent) to 5 (poor) to each of the 16 cases. The distributions of the assigned grades are visualized in Fig. 7, where Ours-S is better than Ours-J as well as Cycle in terms of both quality criteria.

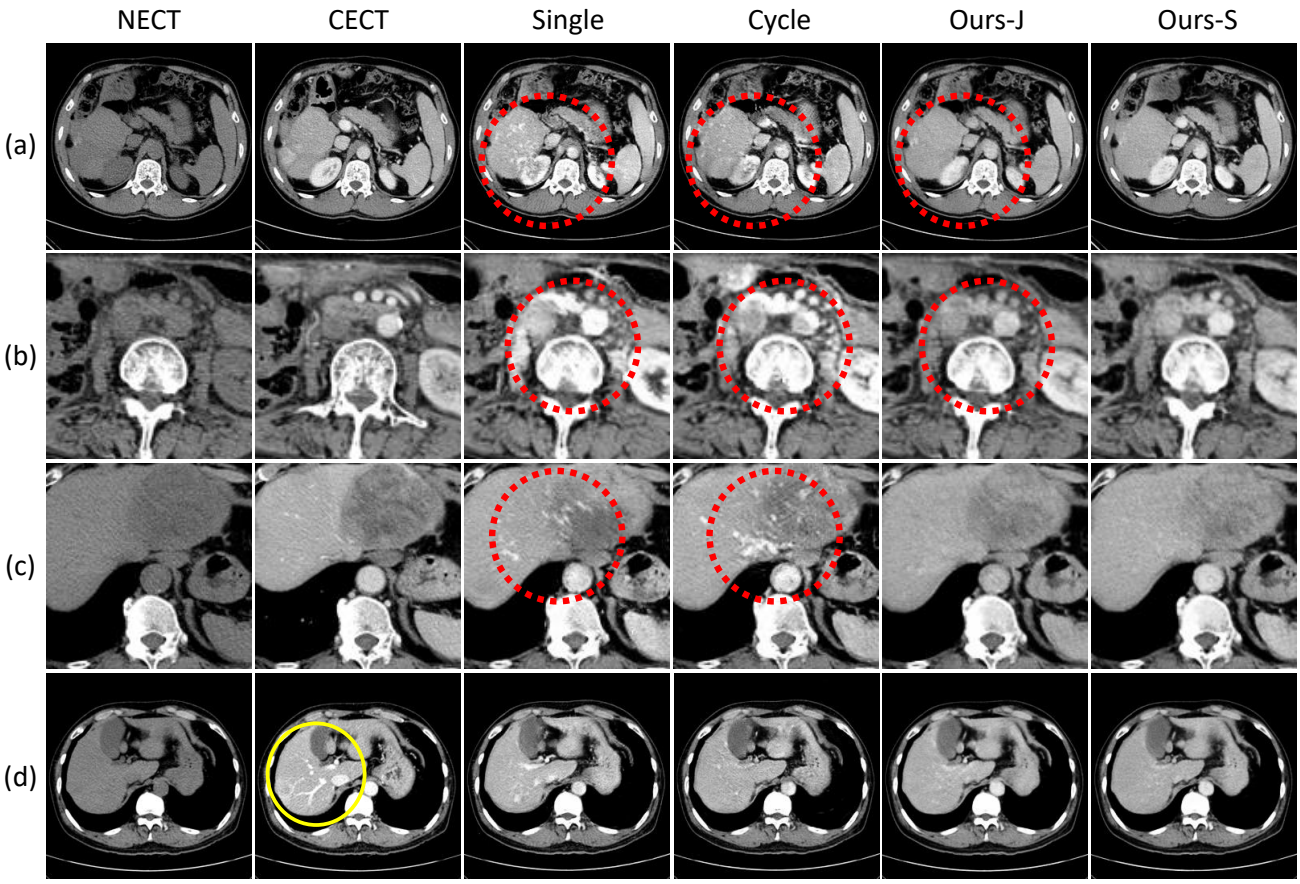

Figure 6. Qualitative results on the abdomen CT image dataset. Red circles indicate artifacts on organs or unclear contrast patterns. Yellow circle indicates the contrast patterns that where all methods fail to synthesize accurately.

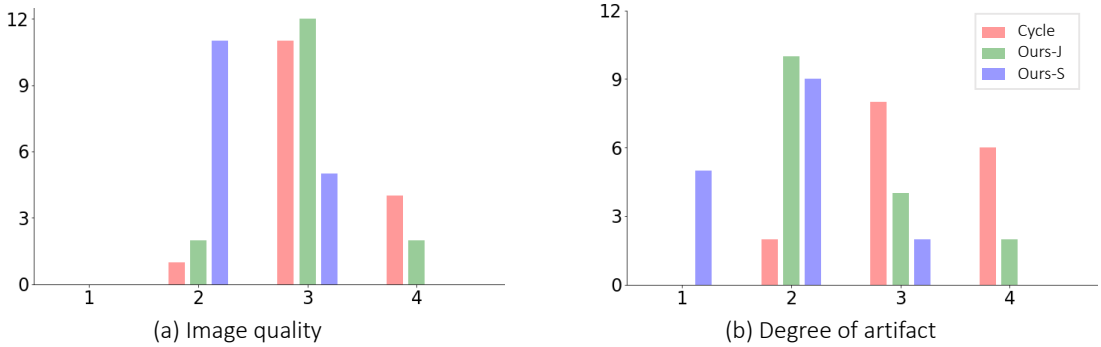

Figure 7. Qualitative comparisons of our model and the baselines by physicians.

4.4. Results on IXI Brain MRI Dataset

4.4.1 Dataset Specification

The IXI brain MRI dataset [1] is a collection of brain MR images taken in several modalities. In particular, T2 (T2-weighted) and PD (proton density) images are aligned perfectly per subject since they are captured simultaneously, unlike CT scans taken sequentially. Thanks to the aligned image pairs, the dataset is appropriate for precisely quantifying the performance of image synthesis models.

Disregarding subjects with only a small number of images, we collect in total 566 pairs of T2 and PD subjects from 566 patients. For the collection, following preprocessing steps are conducted. First, from the image sequence of each subject, only 64 images in the middle are kept for training and evaluation; the others are not used since large

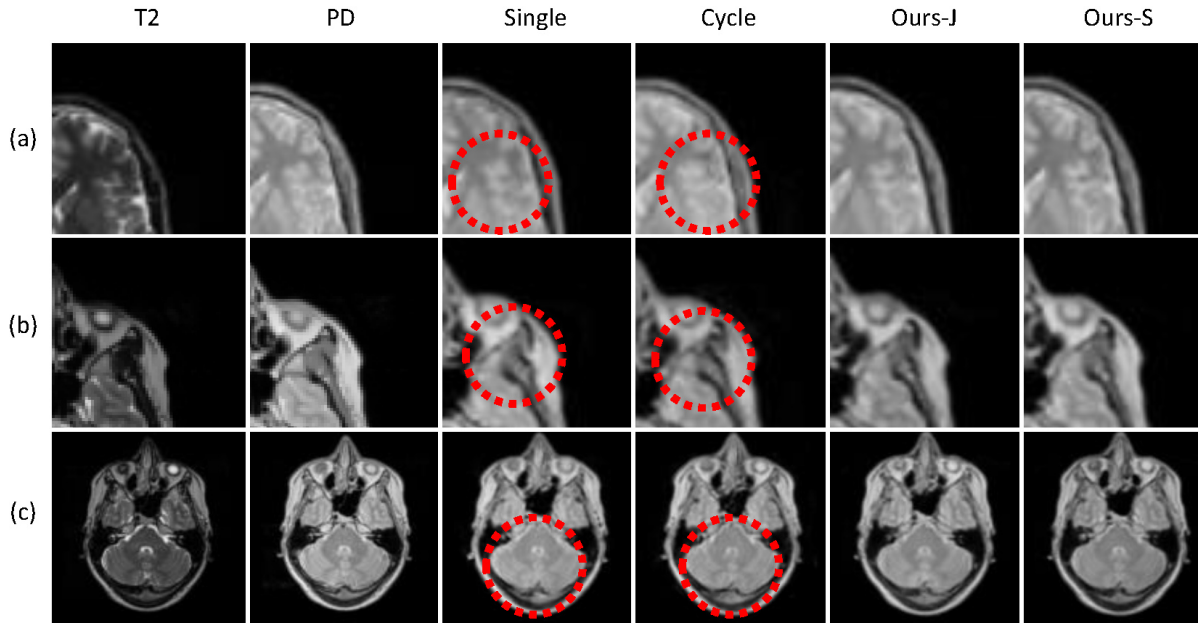

Figure 8. Qualitative results on the IXI brain MRI dataset with $k = 0.05$. Red circles highlight tissues distorted in synthetic PD images.

portions of them are blank areas. We also resize the images to 128×128 to reduce the computational cost.

To evaluate how well our models and the baselines handle misaligned training images, we simulate misalignment between T2 and PD images for training while the evaluation is done with the aligned images of the dataset as is. For this purpose, we apply random affine transformations with various degrees of distortion to the aligned PD images. In detail, an affine transformation matrix is formulated as $\mathbf{I} + \mathbf{k}\,\mathcal{U}$, where $\mathbf{I}$ is 4×4 identity matrix and $\mathcal{U}$ is a matrix of the same size whose elements are uniformly sampled from the interval $[-1, 1]$. Also, k is a value sampled from $\{0.01, 0.02, 0.05\}$; we generate three training sets with the three different values of k to see and analyze the effect of the degree of misalignment quantitatively.

4.4.2 Performance Analysis

Our models and the baseline methods are learned on the training sets with three different degrees of distortion, and evaluated on the perfectly aligned test images. The quantitative results in Tab. 1 show that our models outperform the baselines for all metrics except MSE when k is 0.05. As k increases, the performance of Single drops significantly, while our methods are more robust to the distortion. The performance of Cycle fluctuates within a narrow range, but its overall performance is inferior to that of ours.

In Fig. 8, qualitative results on the IXI brain MRI dataset show similar tendency with those of abdomen CT image dataset. In detail, as shown in Fig. 8, Single and Cycle occasionally synthesize irrelevant patterns, while Ours-J and Ours-S keep the underlying pattern of the input image.

5. Conclusion

We have presented a deep learning framework for synthesizing CECT images given NECT images without using contrast materials. During training our method effectively deals with misalignment between CECT and NECT images by synthesizing well-aligned synthetic NECT images, which enable us to utilize strong reconstruction losses. Experimental results have demonstrated the effectiveness of our method, and its advantages over existing neural image translation techniques have been verified by physicians.

However, the quality of synthetic CECT images given by our method is not accurate enough. Further improvement could be achieved by aligning NECT and CECT images during training; as demonstrated in the IXI dataset, less distorted training images lead to more accurate image synthesis. As a future direction, we thus aim to jointly solve the original task and the registration between real and synthetic NECT imges; this registration task will be easier and can be used to register real NECT and CECT images synthetic NECT is already aligned with real CECT.

Acknowledgement: This work was supported in part by IITP grant and Basic Science Research Program through the NRF funded by the Korea government (MSIT) (No.2019-0-01906 Artificial Intelligence Graduate School Program (POSTECH), 2018R1C1B6001223, NRF-2018R1A5A1060031, 2020-0-00153, 2016-0-00464), and in part by a study on the HPC Support Project supported by MSIT and NIPA.

References

[1] IXI dataset. http://brain-development.org/ixi-dataset. Accessed: 2019-11-13.

[2] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. An unsupervised learning model for deformable medical image registration. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9252–9260, 2018.

[3] J E Barnes. Characteristics and control of contrast in ct. *RadioGraphics*, 12(4):825–837, 1992.

[4] Hu Chen, Yi Zhang, Mannudeep K Kalra, Feng Lin, Yang Chen, Peixi Liao, Jiliu Zhou, and Ge Wang. Low-dose ct with a residual encoder-decoder convolutional neural network. *IEEE transactions on medical imaging*, 36(12):2524–2535, 2017.

[5] Qifeng Chen and Vladlen Koltun. Photographic image synthesis with cascaded refinement networks. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, 2017.

[6] Joseph Paul Cohen, Margaux Luck, and Sina Honari. Distribution matching losses can hallucinate features in medical image translation. In *MICCAI*, pages 529–536. Springer, 2018.

[7] Thomas de Bel, Meyke Hermsen, Jesper Kers, Jeroen van der Laak, and Geert Litjens. Stain-transforming cycle-consistent generative adversarial networks for improved segmentation of renal histopathology. In *Proc. International Conference on Medical Imaging with Deep Learning (MIDL)*, 2019.

[8] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*, 2015.

[9] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[10] Leon A Gatys, Alexander S Ecker, Matthias Bethge, Aaron Hertzmann, and Eli Shechtman. Controlling perceptual factors in neural style transfer. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[11] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *Proc. European Conference on Computer Vision (ECCV)*, 2018.

[12] Yuankai Huo, Zhoubing Xu, Shunxing Bao, Albert Assad, Richard G Abramson, and Bennett A Landman. Adversarial synthesis learning enables segmentation without target modality ground truth. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1217–1220. IEEE, 2018.

[13] Juan Eugenio Iglesias, Ender Konukoglu, Darko Zikic, Ben Glocker, Koen Van Leemput, and Bruce Fischl. Is synthesizing mri contrast useful for inter-modality analysis? In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2013.

[14] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Globally and Locally Consistent Image Completion. In *SIGGRAPH*, 2017.

[15] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proc. International Conference on Machine Learning (ICML)*, 2015.

[16] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1125–1134, 2017.

[17] Hans J Johnson and Gary E Christensen. Consistent landmark and intensity-based image registration. *IEEE transactions on medical imaging*, 21(5):450–461, 2002.

[18] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Proc. European Conference on Computer Vision (ECCV)*, pages 694–711. Springer, 2016.

[19] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. Learning to discover cross-domain relations with generative adversarial networks. In *Proc. International Conference on Machine Learning (ICML)*, 2017.

[20] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proc. International Conference on Learning Representations (ICLR)*, 2015.

[21] Stefan Klein, Marius Staring, Keelin Murphy, Max A Viergever, and Josien PW Pluim. Elastix: a toolbox for intensity-based medical image registration. *IEEE transactions on medical imaging*, 29(1):196–205, 2010.

[22] Chuan Li and Michael Wand. Precomputed real-time texture synthesis with markovian generative adversarial networks. In *Proc. European Conference on Computer Vision (ECCV)*, 2016.

[23] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. In *Proc. Neural Information Processing Systems (NeurIPS)*, pages 700–708, 2017.

[24] Fujun Luan, Sylvain Paris, Eli Shechtman, and Kavita Bala. Deep photo style transfer. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[25] Fujun Luan, Sylvain Paris, Eli Shechtman, and Kavita Bala. Deep painterly harmonization. In *Computer Graphics Forum*, 2018.

[26] Andrew L. Maas, Awni Y. Hannun, and Andrew Y. Ng. Rectifier nonlinearities improve neural network acoustic models. In *ICML Workshop on Deep Learning for Audio, Speech and Language Processing*, 2013.

[27] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[28] Andriy Myronenko and Xubo Song. Intensity-based image registration by minimizing residual complexity. *IEEE transactions on medical imaging*, 29(11):1882–1891, 2010.

[29] Dong Nie, Roger Trullo, Jun Lian, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. Medical image synthesis with context-aware generative adversarial networks. In *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 417–425. Springer, 2017.

[30] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[31] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *AutoDiff, NIPS Workshop*, 2017.

[32] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. In *Proc. International Conference on Learning Representations (ICLR)*, 2016.

[33] Md Mahfuzur Rahman Siddiquee, Zongwei Zhou, Nima Tajbakhsh, Ruibin Feng, Michael B Gotway, Yoshua Bengio, and Jianming Liang. Learning fixed points in generative adversarial networks: From image-to-image translation to disease detection and localization. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, 2019.

[34] Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee. Generative adversarial text to image synthesis. In *Proc. International Conference on Machine Learning (ICML)*, 2016.

[35] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *Proc. International Conference on Learning Representations (ICLR)*, 2015.

[36] Yi-Hsuan Tsai, Xiaohui Shen, Zhe Lin, Kalyan Sunkavalli, Xin Lu, and Ming-Hsuan Yang. Deep image harmonization. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[37] Raviteja Vemulapalli, Hien Van Nguyen, and Shaohua Kevin Zhou. Unsupervised cross-modal synthesis of subject-specific scans. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, 2015.

[38] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *CVPR*, 2018.

[39] Jelmer M Wolterink, Anna M Dinkla, Mark HF Savenije, Peter R Seevinck, Cornelis AT van den Berg, and Ivana Išgum. Deep mr to ct synthesis using unpaired data. In *International Workshop on Simulation and Synthesis in Medical Imaging*, pages 14–23. Springer, 2017.

[40] Tian Xia, Agisilaos Chartsias, and Sotirios A. Tsaftaris. Adversarial pseudo healthy synthesis needs pathology factorization. In *Proc. International Conference on Medical Imaging with Deep Learning (MIDL)*, 2019.

[41] Zizhao Zhang, Lin Yang, and Yefeng Zheng. Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[42] Zhiming Zhou, Han Cai, Shu Rong, Yuxuan Song, Kan Ren, Weinan Zhang, Yong Yu, and Jun Wang. Activation maximization generative adversarial nets. In *Proc. International Conference on Learning Representations (ICLR)*, 2018.

[43] Jun-Yan Zhu, Philipp Krähenbühl, Eli Shechtman, and Alexei A Efros. Generative Visual Manipulation on the Natural Image Manifold. In *Proc. European Conference on Computer Vision (ECCV)*, 2016.

[44] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 2223–2232, 2017.