



Anatomically constrained CT image translation for heterogeneous blood vessel segmentation

Giammarco La Barbera, Haithem Boussaid, Francesco Maso, Sabine Sarnacki, Laurence Rouet, Pietro Gori, Isabelle Bloch

► To cite this version:

Giammarco La Barbera, Haithem Boussaid, Francesco Maso, Sabine Sarnacki, Laurence Rouet, et al.. Anatomically constrained CT image translation for heterogeneous blood vessel segmentation. BMVC 2022 - The 33rd British Machine Vision Conference, Nov 2022, London, United Kingdom. hal-03797472

HAL Id: hal-03797472

<https://telecom-paris.hal.science/hal-03797472>

Submitted on 4 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Anatomically constrained CT image translation for heterogeneous blood vessel segmentation

Giammarco La Barbera¹
 giammarco.labarbera@telecom-paris.fr
 Haithem Boussaid^{*6,2}

¹ LTCI
 Télécom Paris, Institut Polytechnique
 de Paris
 France

Francesco Maso¹

² Philips Research Paris
 Suresnes, France

Sabine Sarnacki^{3,4}

³ IMAG2
 Institut Imagine, Université Paris Cité
 France

Laurence Rouet²

⁴ Université Paris Cité
 Hôpital Necker Enfants-Malades, APHP
 France

Pietro Gori¹

⁵ Sorbonne Université
 CNRS, LIP6
 Paris, France

Isabelle Bloch^{5,1,3}

⁶ Technology Innovation Institute
 Abu Dhabi, UAE

Abstract

Anatomical structures such as blood vessels in contrast-enhanced CT (ceCT) images can be challenging to segment due to the variability in contrast medium diffusion. The combined use of ceCT and contrast-free (CT) CT images can improve the segmentation performances, but at the cost of a double radiation exposure. To limit the radiation dose, generative models could be used to synthesize one modality, instead of acquiring it. The CycleGAN approach has recently attracted particular attention because it alleviates the need for paired data that are difficult to obtain. Despite the great performances demonstrated in the literature, limitations still remain when dealing with 3D volumes generated slice by slice from unpaired datasets with different fields of view. We present an extension of CycleGAN to generate high fidelity images, with good structural consistency, in this context. We leverage anatomical constraints and automatic region of interest selection by adapting the *Self-Supervised Body Regressor*. These constraints enforce anatomical consistency and allow feeding anatomically-paired input images to the algorithm. Results show qualitative and quantitative improvements, compared to state-of-the-art methods, on the translation task between ceCT and CT images (and vice versa).

Moreover, using the CT images produced by our algorithm, we achieve blood vessel segmentation performance on par with the segmentation performance using real CT images.

1 Introduction

Heterogeneity in contrast is one of the major difficulties in medical image segmentation when using Convolutional Neural Networks (CNN), in particular in contrast-enhanced Computed Tomography (ceCT) images. The effect of the contrast agent on the pixel intensity is not always the same among patients due to different factors, such as acquisition times and patient morphology. Furthermore, the presence of a tumor or thrombosis in the vessels can also cause heterogeneity in contrast within an anatomical structure. This raises difficulties during segmentation, and manual corrections are often needed.

In [23, 27, 53] the authors show that the combined use of ceCT and contrast-free (CT) CT images is able to deal with the heterogeneity of ceCT images and thus improves segmentation. However, in order to limit ionising radiations, clinicians often acquire only one CT modality. One common computational approach to compensate for the absence of an imaging modality is to use generative models [9, 51] to synthesise it. In the absence of paired data sets, unsupervised translation methods, based on CycleGAN [52] and UNIT [17], have been proposed [9, 14, 21, 51]. Some authors have also already considered applying CycleGAN [23, 27] or UNIT [53] to artificially remove or add contrast medium on CT images. CycleGAN [52] is an evolution of Generative Adversarial Network (GAN) [10], which introduces a second neural network that tries to solve the inverse task, namely reconstructing the input. A cycle consistency loss function is combined to the adversarial loss to overcome the lack of paired data. UNIT [17] is another model conceived for the unpaired setting. This generative model is composed of two variational autoencoder networks, which work on two different domains but share the same latent space. Different modifications have already been proposed for both methods, such as the use of Wasserstein distance [9], attention mechanisms [8, 5] and U-Net as discriminator network [24]. However, these models do not guarantee to preserve fine structures [53] and may produce artefacts [25, 51], which prevent their use for the segmentation of small and heterogeneous structures, such as blood vessels. In particular, the cycle consistency loss function enforces a relationship only at a distribution level. In [19], the authors demonstrate that CycleGAN can deliver ambiguous solutions, especially for substantially different distributions as in medical imaging. Several works tried to address this limitation by adding more terms to the loss function, such as mutual information [9] and perceptual loss term [9, 29], that require no supervision. Despite the different methods proposed, the anatomical constraint remains insufficient.

As a matter of fact, another challenge when dealing with unpaired 3D medical images is the lack of 3D consistency. With current hardware memory limitations, it is difficult to train a 3D network taking as input a whole 3D volume. Instead, a common approach is to use 2D networks that take a slice of a 3D volume along one axis. Moreover, in the unpaired scenario we can have different number of slices for the same anatomical region among patients, leading to difficulties to select anatomically-paired slices. In fact, in [25], authors showed that it is fundamental to inform the generator on the specific regions that should be affected by the contrast materials. For this reason, the use of *not-aligned* paired data is more effective than unpaired data. Some authors [50] claim that the use of unpaired data can be mitigated by exploiting the approximately common anatomy between subjects. They refer to this as *position-based selection (PBS)* strategy. However, in the abdominal

region, the different sizes and lengths of the organs must be taken into account, implying that the slice n of the patient i with N slices may not have the same anatomical content as slice $m = n \cdot \frac{M}{N}$ of patient j with M slices. Eventually, the use of 3D *affine* registration (e.g., as in the Simple-Elastix [18] library) could be a solution to the problem, but the difference between the two domains, the difficulty of identifying the fixed reference image, and the high variability in shape and relative size and pose of abdominal organs among subjects (especially in 3D) may lead to misalignment.

To address these issues, we propose an extension of the CycleGAN method which includes:

- (i) the automatic selection of the region of interest by exploiting anatomical information, in order to reduce the anatomical distribution of 3D data acquired with different fields of view;
- (ii) the use of a *Self-Supervised Body Regressor* (SSBR), adapted from [28], to select anatomically-paired slices among the unpaired ceCT and CT domains, and help the discriminator to specialize in the task;
- (iii) the use of the SSBR score as an extra loss function that constrains the generator to produce a slice describing the same anatomical content as the input, inspired from the auxiliary classifier GAN [29];
- (iv) the use of the input image as a template for the generator, as in [9], and the use of an anatomical binary mask to constrain the output.

The proposed method is generic and could be used in different medical applications, i.e. different body regions such as brain or lungs, or different translation modalities such as MRI to CT or T1-w to T2-w. Here, we propose to use it for the generation of CT abdominal images from ceCT images and vice versa. We test the use of a generated modality, in combination with the complementary original one, to improve segmentation performance on blood vessels of pathological patients. To the best of our knowledge, this strategy has never been tested for such an application.

We show that our method greatly improves the ceCT-CT translation quality compared to state-of-the-art methods. As a consequence, the segmentation performances using generated images are also improved, achieving both qualitative and quantitative results comparable to the ones using both real images.

It is important to highlight that, in this work, the use of synthetic images is intended to increase segmentation performances and not to use for clinical diagnosis.

It should be also noted that in this paper we focus on CNN-based methods, as state-of-the-art methods for unsupervised medical image translation. While interesting works on image-to-image translation based on *transformers* are starting to be explored [34, 35], their application in the medical domain is limited due to the restricted number of data available. For this reason, existing works focus only on paired medical data sets [6]. Nevertheless, for the sake of completeness, we test a transformer-GAN, namely TransGAN [36], on our unsupervised medical task.

2 Proposed Method

As stated in Section 1, when using 2D unpaired medical data, the selection of consistent (i.e., corresponding to the same region of interest (ROI)) and anatomically similar slices between the two domains is highly important to facilitate the generative process. To this end,

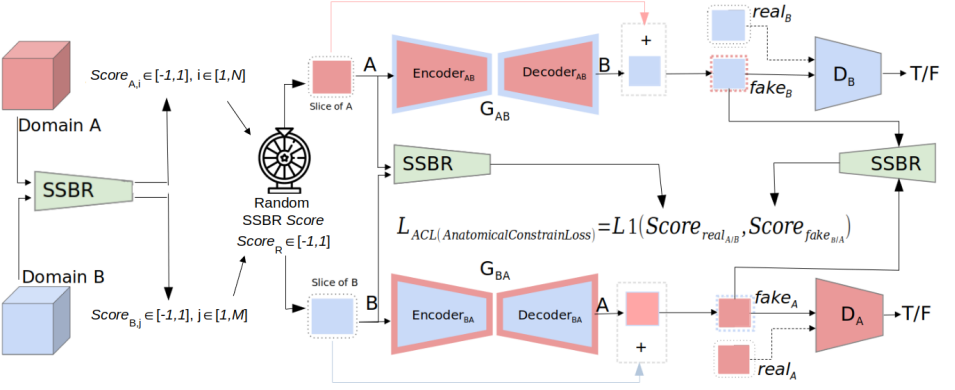


Figure 1: Proposed method for the selection of anatomically-paired slices via *Self-Supervised Body Regressor*, and its use as a loss function L_{ACL} .

we propose to leverage a Self-Supervised Body Regressor (SSBR) [28], a CNN that finds common features on anatomically similar slices from unlabeled CT images. This results in assigning the same label for slices describing the same anatomy while belonging to different patients. The SSBR is trained to estimate slice scores which are monotonous functions of the slice indices. However, there is no guarantee to obtain the same range of scores for different modalities. We propose a solution to this problem. The method described in this section is summarized in Figure 1.

Input selection via SSBR First of all, because of the different fields of view (FOV) in the two datasets, it is important to select an appropriate ROI. This can be done off-line and manually, as in [23, 33]. Here, instead, we first propose a simple, automatic and on-line method to select only slices from the abdominal region. We automatically select the first slice of the lungs and the last slice of the intestinal area as upper and lower landmarks, which is easy due to the strong presence of black pixels in both ceCT and CT acquisitions.

Then, instead than PBS strategy as in [30], we propose the use of an SSBR, as shown in Figure 1. For training, we optimize three loss functions that do not require annotated anatomical labels. The first one, as in [28], favors an increasing order of SSBR scores according to the positions of the slices, avoiding repeating scores and ensuring similar scores for adjacent regions:

$$L_{order} = - \sum_{k=1}^K \sum_{p=1}^{P-1} \log(h(\text{Score}_{k,p+1} - \text{Score}_{k,p})) \quad (1)$$

where $\text{Score}_{k,p} \in [-1, 1]$ is the SSBR output for slice p of CT volume k , h is the sigmoid activation function, K is the number of CT volumes in the chosen set (mini-batch) and P is the number of slices in each volume.

The second loss function exploits the automatic selection of the ROI, forcing the first and last slices to have a score of -1 and +1 respectively:

$$L_{norm} = \sum_{k=1}^K (f(\text{Score}_{k,1} + 1) + f(\text{Score}_{k,P} - 1)) \quad (2)$$

where f is a smoothed L1 norm. This function guarantees the same score range for both modalities.

The third loss function takes into account the anatomical variability of the abdominal area. Using the binary mask BM of the body for each slice (easily obtained in CT), we want the difference between successive scores to be an increasing function of the normalized cardinality of the intersection of the BM of successive slices:

$$L_{anat} = \sum_{k=1}^K \sum_{p=1}^{N-1} f(\Delta_{k,p+1}^{BM} - \Delta_{k,p}^{BM}) \quad (3)$$

with $\Delta_{k,p}^{BM} = 1 - \frac{|BM_{k,p} \cap BM_{k,p-1}|}{|BM_{k,p-1}|}$ and $\Delta_{k,p} = Score_{k,p} - Score_{k,p-1}$

This is done in order to increase the difference in score between slices with higher anatomical difference and not fall into the trivial linear solution.

Eventually, the terms of the cost function are combined by a weighted average, and the function to be optimized is:

$$L_{SSBR} = \alpha L_{order} + \beta L_{anat} + \gamma L_{norm} \quad (4)$$

where α , β and γ are empirically chosen weights that balance the three losses.

Once the SSBR is properly trained (details in the next section), to extract the anatomically-paired slices for each iteration of the CycleGAN we do the following:

1. A single patient is selected for each of the unpaired ceCT and CT domains, called domains A and B;
2. The 3D volumes are automatically restricted to the abdominal region;
3. SSBR scores are predicted for each 2D slice of the two 3D ROIs, using the pre-trained SSBR;
4. J random SSBR scores, denoted by $Score_{R_j}$, are sampled in $[-1, 1]$, where J is the selected number of slices corresponding to the size of the mini-batch;
5. For each $Score_{R_j}$, the slice with the closest score is selected in each domain, as $\arg \min_p |Score_{R_j} - Score_{\cdot,p}|$ where \cdot is the domain (A or B) and p is the selected slice in $[1, N]$ for A and $[1, M]$ for B.

Anatomically constrained CycleGAN Inspired by [20], we propose the use of the pre-trained SSBR as an auxiliary classifier to enforce the anatomical consistency (i.e., same body parts) between the input and the synthesized output. During the training phase of the generator, we add to the loss functions of the standard CycleGAN an L1 norm between the SSBR score of the input *real* A (resp. *real* B) and the SSBR score of the generated slice *fake* B (resp. *fake* A), called the *Anatomical Constraint Loss (ACL)*, as shown in Figure 1:

$$L_{ACL} = \frac{1}{J} \sum_{j=1}^J |Score_{real_{A/B,j}} - Score_{fake_{B/A,j}}| \quad (5)$$

We also propose to further constrain the models in two ways. First, as in [8], we use the input image as a template (I_{in} in images and tables), i.e. the generators only need to estimate how to modify the input image without estimating an output image from scratch. Secondly, during inference, we remove the artefacts created in the original black areas (e.g. background or air). Here, we use the binary mask BM used also in Equation 3.

3 Results and Discussion

3.1 Implementation details

The hyperparameters for CycleGAN and UNIT were found empirically on the training set, starting from those in [32] and [14] respectively. The best combination of weights for CycleGAN losses was found as 0.5 for identity loss, 10 for cycle-consistency loss and 1 for both adversarial loss and our L_{ACL} loss function. For KL loss of UNIT we set a weight of 0.01.

For the SSBR, we operated as in [28], using ResNet-34 [14] as the backbone. The best weights in Eq. 4 were found empirically as: $\alpha = 5 \cdot 10^{-3}$, $\beta = 1$, $\gamma = 10$.

All trainings and tests were run on a GPU NVIDIA® Tesla® P100 with 16 GB of VRAM using a mini-batch size of 8.

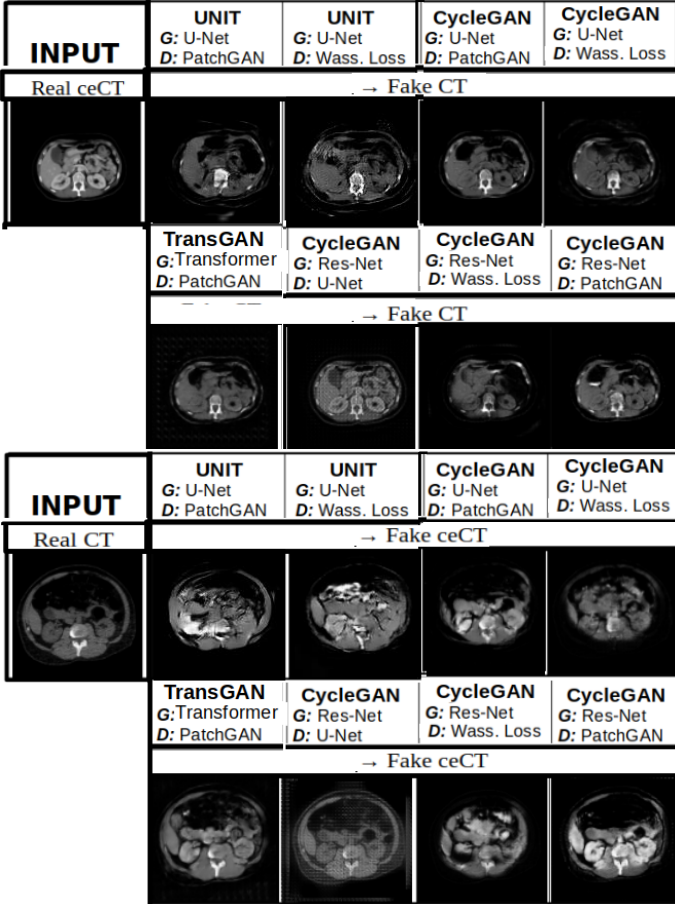


Figure 2: Comparison of some state-of-the-art methods on slices of the unpaired test set. Top table: ceCT to CT. Bottom table: CT to ceCT. The slices in all tests are selected with PBS. The input in the other direction gives an idea of what the expected result should look like.

3.2 Qualitative results on unpaired datasets

Dataset For the unpaired image-to-image translation training, the images were obtained from the Cancer Imaging Archive (TCIA) [5]. Two databases of pathology-free abdominal CT images were used: *CT Pancreas* [24] with 82 ceCT images of abdomens, and *CT Colonography* [26] which contains non-contrast CT images of which 82 healthy subjects were retained for consistency with the first dataset. For both datasets, 72 patients are used for training and 10 for testing. All axial slices were resized from 512×512 to 128×128 pixels to make the training less computationally and memory intensive.

Experiments First, we tested several existing methods [0, 8, 17, 32] to select the best networks for the generators (G) and the discriminators (D). The use of the only renal region slices was deemed essential for our experiments to obtain anatomically consistent images and our automatic detection of the abdominal region proved effective, removing the need for manual selection. Some qualitative examples about automatic ROI selection is showed in supplementary materials. Moreover, in all these tests, we used the PBS [30] strategy for selecting slices at the same relative position. Although different methods gave satisfactory results for the easier task of ceCT2CT, only the CycleGAN [32] with ResNet as the generating network and PatchGAN as the discriminating mechanism produced good results in terms of contrast realness for the task of CT2ceCT, as shown in Figure 2. Probably for methods such as TransGAN, performances are limited by the restricted amount of data and computational power.

Despite the good results shown in the method identified as the best ones in Figure 2, in terms of overall shape and contrast intensity, the PBS selection was not sufficient and several anatomical artefacts appeared (see Figure 3). Another existing strategy for anatomically-paired selection that we tested was the use of 3D *affine* registration with Simple-Elastix [18] algorithm. Given the high variability between the two domains, we decided to perform the registration at each iteration between the two selected patients. Anatomical coherence was improved but some important artefacts still appeared. Finally, our proposed selection with SSBR was tested, which reduced the severity of artefacts, as shown in the forth column of Figure 3.

We then added the L_{ACL} loss function, which significantly improved anatomical coherence, particularly in the binary mask regions. Eventually, we combined in a first moment the use of input as template $I_n A_d$ and in a second moment the binary mask BM . The complete proposed method based on SSBR selection with L_{ACL} , $I_n A_d$ and BM produced high quality synthetic images, without visual artefacts and with realistic contrast intensity according to physicians' evaluation. Some qualitative results are detailed in Figure 3.

3.3 Quantitative ablation study on paired database

Dataset For quantitative testing we used a pathological private pediatric database of paired abdominal ceCT-CT images of 10 patients with renal cancer. It is important to note that the small number of patients in this data set prevents achieving satisfactory performance on training generative models. No public paired ceCT-CT datasets is available, so this small dataset we gathered is quite rare. Moreover, we want to emphasize that for each subject we have at our disposal about 100 2D slices of the renal ROI and therefore the results refer to a total of about 1000 2D images.

Results The quantitative ablation and comparative study was performed using the presented methods, pre-trained on the unpaired data-sets. The results are presented in Table 1 using

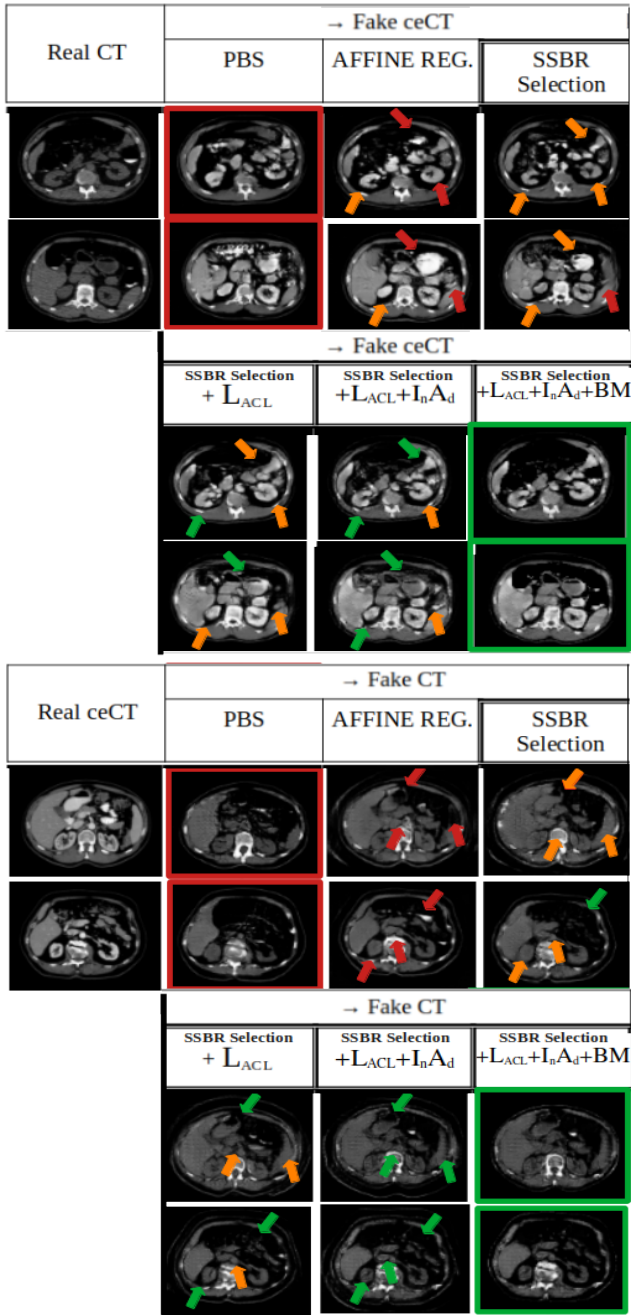


Figure 3: Qualitative results on unpaired slices. Top table: CT to ceCT. Bottom table: ceCT to CT. Tests based on CycleGAN. Last four columns: results with our methods, we add each proposition to the SSBR selection. $I_r A_d$ indicates the input addition while BM the use of binary mask. Arrows: high (red), low (orange) and no (green) artefacts.

mean square error (MSE), structure similarity (SSIM) and peak signal to noise ratio (PSNR) between *real* and *fake* images, and training time (TIME). All our contributions improve on the original CycleGAN with PBS [50], at the cost of some additional learning time (note that the inference time remains the same for all methods). Their combination produces the best results for both tasks. The use of *affine* registration seems to be quantitatively comparable to the use of SSBR for the selection and loss function L_{ACL} , but the network requires a very high computational time in addition to the creation of artefacts, as illustrated in Figure 3.

Table 1: Quantitative study on the 10 patients with paired images. Mean square error (MSE), structure similarity (SSIM) and peak signal to noise ratio (PSNR) are shown as mean and standard deviation. TIME is the training time. Tests based on CycleGAN. The first section of rows presents the ablation study, in which each proposition is added to the SSBR selection. I_nA_d indicates the input addition while BM the use of binary mask.

CycleGAN Method	MSE [10^{-2}] (↓)	SSIM [10^{-1}] (↑)	PSNR (↑)	TIME (↓)
real CT→fake ceCT vs real ceCT				
$+L_{ACL} + I_nA_d + BM$	6.37 (2.01)	6.81 (0.62)	18.14 (1.23)	7h 55m
$+L_{ACL} + I_nA_d$	6.41 (1.97)	6.67 (0.63)	18.11 (1.22)	7h 55m
$+L_{ACL} + BM$	8.19 (2.32)	6.36 (0.72)	17.02 (1.14)	7h 49m
$+I_nA_d$	6.79 (2.85)	6.60 (0.74)	17.97 (1.54)	7h 14m
$+BM$	8.42 (2.46)	6.24 (0.73)	16.91 (1.17)	7h 5m
$+L_{ACL}$	8.55 (2.28)	6.19 (0.69)	16.82 (1.07)	7h 49m
SSBR selection	9.07 (2.39)	5.99 (0.71)	16.56 (1.07)	7h 5m
AFFINE REG.	8.16 (1.80)	6.36 (0.57)	16.99 (0.87)	16h 33m
PBS	10.05 (2.89)	5.76 (0.65)	16.14 (1.15)	3h 2m
real ceCT→fake CT vs real CT				
$+L_{ACL} + I_nA_d + BM$	4.05 (0.83)	7.23 (0.53)	20.03 (0.92)	7h 55m
$+L_{ACL} + I_nA_d$	4.24 (0.86)	6.80 (0.37)	19.83 (0.92)	7h 55m
$+L_{ACL} + BM$	5.08 (0.85)	6.87 (0.52)	19.02 (0.74)	7h 49m
$+I_nA_d$	6.16 (1.15)	5.87 (0.23)	18.18 (0.79)	7h 14m
$+BM$	6.07 (1.28)	6.61 (0.65)	18.28 (0.99)	7h 5m
$+L_{ACL}$	5.87 (1.73)	6.08 (0.22)	18.47 (1.12)	7h 49m
SSBR selection	7.15 (2.16)	5.68 (0.52)	17.64 (1.26)	7h 5m
AFFINE REG.	4.72 (0.95)	6.77 (0.37)	19.36 (0.93)	16h 33m
PBS	8.26 (1.97)	5.36 (0.28)	16.96 (1.04)	3h 2m

3.4 Blood vessel segmentation using ceCT and CT

Dataset For the proposed segmentation application, the synthetic images used were produced using generative methods trained as explained previously but with images at the original size 512×512 . Reference segmentations of arteries and veins were manually performed by medical experts on our paired pathological dataset.

Segmentation performances To further demonstrate the realness of the images generated by our method, similarly to [9, 23, 27, 33], we compared the performance of a segmentation network when using either a real image and a fake image, or both real images. Given the restricted dataset, all tests were done with the Leave-One-Patient-Out (L-O-P-O) method using the 3D nnU-Net [22]. Results show that replacing a real CT modality with a synthetic one produced with CycleGAN and the PBS method, as in [23, 27], is not sufficient to achieve performances as good as when using both real modalities. By contrast, the synthetic CT images produced by our method achieve the highest Dice score and the lowest Hausdorff distance, with the best combination of precision and recall. This is even more evident for

the more heterogeneous cases, particularly for the veins. Quantitative results are shown in Table 2. Some qualitative results and other quantitative results for an extended ceCT dataset (no CT images available) can be found in the supplementary material.

Table 2: Segmentation performance on **real ceCT** of 10 patients (and then on the only 5 more heterogeneous cases) using L-O-P-O methods. Dice score (DS), precision (PR), recall (RC) and 95th percentile of the Hausdorff distance (HD95) are given (mean and standard deviation). All tests were done using 3D nnU-Net [16] with intensity (except if indicated) and geometric data augmentation.

INPUT Database	Structure	DS [100%] (↑)	PR [100%] (↑)	RC [100%] (↑)	HD95 [mm] (↓)
on 10 patients					
real ceCT and real CT	Arteries	74.61 (5.89)	85.22 (8.32)	69.06 (8.15)	15.39 (5.72)
	Veins	45.62 (13.72)	60.61 (19.53)	38.68 (14.83)	31.47 (16.53)
real ceCT without data aug.	Arteries	63.75 (11.18)	80.33 (10.99)	53.88 (12.48)	23.43 (8.18)
	Veins	21.18 (19.70)	64.04 (34.08)	15.45 (16.04)	42.14 (23.79)
real ceCT	Arteries	73.01 (6.57)	81.08 (8.70)	67.19 (8.43)	15.80 (7.01)
	Veins	40.58 (23.50)	55.94 (31.39)	33.72 (26.61)	40.65 (30.90)
real ceCT and fake _{PBS} CT	Arteries	69.59 (8.89)	79.54 (10.85)	63.47 (12.59)	18.08 (8.21)
	Veins	44.40 (22.75)	58.44 (21.78)	38.38 (23.20)	39.31 (16.79)
real ceCT and fake _{Ours} CT	Arteries	72.33 (7.41)	77.29 (10.32)	68.63 (8.88)	15.48 (6.38)
	Veins	44.49 (22.50)	54.98 (26.74)	40.28 (22.69)	38.90 (32.76)
on 5 more heterogeneous					
real ceCT and real CT	Arteries	75.01 (5.82)	85.17 (4.37)	67.50 (8.57)	12.79 (6.04)
	Veins	40.87 (14.73)	56.93 (18.63)	32.62 (13.05)	31.16 (10.76)
real ceCT without data aug.	Arteries	66.59 (8.31)	86.89 (5.70)	54.83 (10.29)	23.34 (9.14)
	Veins	14.66 (17.05)	71.31 (39.90)	8.89 (10.98)	50.35 (29.50)
real ceCT	Arteries	72.94 (6.30)	84.37 (3.80)	64.89 (9.71)	13.49 (5.14)
	Veins	28.28 (19.84)	51.97 (38.06)	17.50 (18.41)	35.57 (14.33)
real ceCT and fake _{PBS} CT	Arteries	70.77 (9.18)	84.41 (5.96)	63.00 (15.51)	13.83 (5.95)
	Veins	33.47 (26.92)	45.48 (34.33)	27.73 (23.78)	37.73 (23.42)
real ceCT and fake _{Ours} CT	Arteries	73.18 (7.51)	80.58 (4.59)	67.63 (11.25)	12.73 (4.10)
	Veins	40.57 (20.25)	62.01 (13.31)	31.96 (18.91)	32.83 (13.84)

4 Conclusion

We presented an extension of CycleGAN via the use of a Self-Supervised Body Regressor to: (i) better select anatomically-paired slices; (ii) anatomically constrain the generator to produce a slice describing the same anatomical content as the input. We applied our method to the unsupervised synthesis of ceCT-CT images. We showed significant improvements in the generated images compared to existing methods. To further validate our method, we demonstrated that the synthesized images can be used to guide a segmentation method by compensating, without loss of performance, for the absence of the complementary real acquisition modality. Future work aims to apply our method on other translation tasks, such as MRI to CT or T1-w to T2-w, and other body sections. Moreover, once more data and more powerful GPUs will be available, *transformer*-based methods will be further explored.

Acknowledgments This work has been partially funded by a grant from Region Ile de France (DIM RFSI).

References

- [1] M Arjovsky, S Chintala, and L Bottou. Wasserstein GAN. *ICML*, 70:214–223, 2017.
- [2] K Armanious, C Jiang, S Abdulatif, et al. Unsupervised medical image translation using cycle-medgan. In *European Signal Processing Conference (EUSIPCO)*, pages 1–5, 2019.
- [3] H Chang, J Lu, F Yu, et al. PairedCycleGAN: Asymmetric style transfer for applying and removing makeup. In *IEEE CVPR*, page 40–48, 2018.
- [4] X Chen, C Lian, L Wang, et al. Diverse data augmentation for learning image segmentation with cross-modality annotations. *Medical Image Analysis*, 71:102060, 2021.
- [5] K Clark, B Vendt, K Smith, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *Journal of Digital Imaging*, 26(6):1045–1057, 2013.
- [6] Onat Dalmaz, Mahmut Yurt, and Tolga Cukur. ResViT: Residual vision transformers for multi-modal medical image synthesis. *IEEE TMI*, PP:1–1, 2022.
- [7] S UH Dar, M Yurt, L Karacan, et al. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *IEEE TMI*, 38(10):2375–2388, 2019.
- [8] H Fukui, T Hirakawa, T Yamashita, and H Fujiyoshi. Attention branch network: Learning of attention mechanism for visual explanation. In *IEEE CVPR*, pages 10697–10706, 2019.
- [9] Y Ge, D Wei, Z Xue, et al. Unpaired MR to CT synthesis with explicit structural constrained adversarial learning. In *IEEE ISBI*, 2019.
- [10] I J Goodfellow, J Pouget-Abadie, M Mirza, et al. Generative Adversarial Nets. In *NeurIPS*, page 2672–2680, 2014.
- [11] K He, X Zhang, S Ren, et al. Deep residual learning for image recognition. In *IEEE CVPR*, pages 770–778, 2016.
- [12] F Isensee, P Jaeger, S Kohl, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18:1–9, 2021.
- [13] Yifan Jiang, Shiyu Chang, and Zhangyang Wang. TransGAN: Two pure transformers can make one strong GAN, and that can scale up. *Advances in NIPS*, 34, 2021.
- [14] S Karthika and M Durgadevi. Generative adversarial network (GAN): a general review on different variants of GAN and applications. In *ICCES*, pages 1–8, 2021.
- [15] J Kim, M Kim, H Kang, et al. U-GAT-IT: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. In *ICLR*, pages 1–19, 2020.
- [16] S Kim, Baek, Park J, et al. InstaFormer: Instance-aware image-to-image translation with transformer. In *CVPR*, 2022.

- [17] M Y Liu, T Breuel, and J Kautz. Unsupervised image-to-image translation networks. In *NeurIPS*, pages 700–708, 2017.
- [18] K Marstal, F Berendsen, M Staring, et al. SimpleElastix: A user-friendly, multi-lingual library for medical image registration. In *WBIR*, 2016.
- [19] N Moriakov, J Adler, and J Teuwen. Kernel of CycleGAN as a principle homogeneous space. In *ICLR*, 2020.
- [20] A Odena, C Olah, and J Shlens. Conditional image synthesis with auxiliary classifier GANs. In *ICLR*, 2017.
- [21] Y Pan, M Liu, C Lian, et al. Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer’s disease diagnosis. In *MICCAI*, pages 455–463, 2018.
- [22] H R Roth, A Farag, E B Turkbey, et al. Data from Pancreas-CT - The Cancer Imaging Archive, 2016.
- [23] V Sandfort, K Yan, P Pickhardt, et al. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Scientific Reports*, 9(1):1–9, 2019.
- [24] E Schonfeld, B Schiele, and A Khoreva. A U-Net based discriminator for generative adversarial networks. In *IEEE CVPR*, pages 8207–8216, 2020.
- [25] M Seo, D Kim, K Lee, et al. Neural contrast enhancement of CT image. In *IEEE WACV*, pages 3972–3981, 2021.
- [26] K Smith, K Clark, W Bennett, et al. Data from CT-COLONOGRAPHY - The Cancer Imaging Archive, 2015.
- [27] C Song, B He, H Chen, et al. Non-contrast CT liver segmentation using CycleGAN data augmentation from contrast enhanced CT. In *iMIMIC Workshop (MICCAI)*, pages 122–129, 2020.
- [28] K Yan, L Lu, and R Summers. Unsupervised body part regression via spatially self-ordering convolutional neural networks. In *IEEE ISBI*, pages 1022–1025, 2018.
- [29] C Yang, T Kim, R Wang, et al. ESTHER: Extremely Simple Image Translation Through Self-Regularization. In *BMVC*, 2018.
- [30] H Yang, J Sun, An Carass, et al. Unsupervised MR-to-CT synthesis using structure-constrained CycleGAN. *IEEE TMI*, 39(12):4249–4261, 2020.
- [31] X Yi, E Walia, and P Babyn. Generative adversarial network in medical imaging: A review. *Medical Image Analysis*, 58:101552, 2019.
- [32] J Y Zhu, T Park, P Isola, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *IEEE ICCV*, pages 2223–2232, 2017.
- [33] Y Zhu, Y Tang, Y Tang, et al. Cross-domain medical image translation by shared latent Gaussian mixture model. In *MICCAI*, pages 379–389, 2020.