

Generative Adversarial Networks for MR-CT Deformable Image Registration

Christine Tanner, Firat Ozdemir, Romy Profanter, Valeriy Vishnevsky,
Ender Konukoglu, and Orcun Goksel

Computer Vision Lab, ETH Zurich, 8092 Zurich, Switzerland

Abstract. Deformable Image Registration (DIR) of MR and CT images is one of the most challenging registration task, due to the inherent structural differences of the modalities and the missing dense ground truth. Recently cycle Generative Adversarial Networks (cycle-GANs) have been used to learn the intensity relationship between these 2 modalities for unpaired brain data. Yet its usefulness for DIR was not assessed.

In this study we evaluate the DIR performance for thoracic and abdominal organs after synthesis by cycle-GAN. We show that geometric changes, which differentiate the two populations (e.g. inhale vs. exhale), are readily synthesized as well. This causes substantial problems for any application which relies on spatial correspondences being preserved between the real and the synthesized image (e.g. plan, segmentation, landmark propagation). To alleviate this problem, we investigated reducing the spatial information provided to the discriminator by decreasing the size of its receptive fields.

Image synthesis was learned from 17 unpaired subjects per modality. Registration performance was evaluated with respect to manual segmentations of 11 structures for 3 subjects from the VISERAL challenge. State-of-the-art DIR methods based on Normalized Mutual Information (NMI), Modality Independent Neighborhood Descriptor (MIND) and their novel combination achieved a mean segmentation overlap ratio of 76.7, 67.7, 76.9%, respectively. This dropped to 69.1% or less when registering images synthesized by cycle-GAN based on local correlation, due to the poor performance on the thoracic region, where large lung volume changes were synthesized. Performance for the abdominal region was similar to that of CT-MRI NMI registration (77.4 vs. 78.8%) when using 3D synthesizing MRIs (12 slices) and medium sized receptive fields for the discriminator.

1 Introduction

Deformable Image Registration (DIR) is a challenging task and active field of research in medical image analysis [1]. Its main application is fusion of the information acquired by the different modalities to facilitate diagnosis and treatment planning [1]. For example, in radiotherapy treatment planning Magnet Resonance (MR) images are used to segment the tumor and organs at risk, while the tissue density information provided by the corresponding Computer Tomography

(CT) image is used for dose planning [2]. CT and MR images are acquired using separate devices and often on different days. Therefore the patient will not be in exactly the same posture and the position of inner organs might change, due to respiration, peristalsis, bladder filling, gravity, etc. Thus, DIR is needed. The main difficulty of MR-CT DIR is the definition of an image similarity measure, which reliably quantifies the local image alignment for optimizing the many free parameters of the spatial transformation. This is an inherent problem as multi-modal images are acquired because they provide complementary information.

Multi-modal similarity measures. The main voxel-wise multi-modal image (dis)similarity measures are (i) statistical measures that use intensity information directly and try to maximize (non-linear) statistical dependencies between the intensities of the images (e.g. Normalized Mutual Information (NMI) [3], MI [4]), and (ii) structural measures based on structural representations that try to be invariant to different modalities (e.g. normalized gradient fields [5], entropy images [6], Modality Independent Neighborhood Descriptor (MIND) [7]).

Intensity remapping. The drawback of structural representations is that all unstructured (e.g. homogenous) regions are mapped to the same representation regardless of their original intensity. To avoid this information reduction, methods to directly re-map intensities have been proposed [8,6,9]. The joint histogram of the coarsely registered images was employed to remap the intensities of both images to a common modality based on the least conditional variance to remove structures not visible in both images [8]. Assuming that the global self-similarity of the images (i.e. the similarities between all image patches) is preserved across modalities, intensities were mapped into a 1D Laplacian Eigenmap based on patch intensity similarity [6]. A k-means clustering based binning scheme, to remap spatially unconnected components with similar intensities to distinct intensities, was proposed in [9] for retina images. While these intensity-remappings provide some improvements, they are simplifications to the underlying complex relationship between the intensity of the two modalities.

Learning from paired data. Given aligned multi-modal training data, attempts have been made to learn this complex relationship. The last layer of a deep neural network (DNN) classifier, which discriminates between matching and not matching patches, was used to directly learn the similarity measure [10]. The DNN was initialized by a stacked denoised autoencoder, where the lower layers were separately trained per modality to get modality-dependent features. It was observed that the learned CT filters look mostly like edge-detectors, while the MR filters detect more complex texture features. In [11] the expected joint intensity distribution was learned from co-registered images. The dissimilarity measure was then based on the Bhattacharyya distance between the expected and observed distribution. Machine learning has been used to learn how to map one modality to the other. [12] synthesized CT from MR brain images by matching MR patches to an atlas (created from co-registered MR and CT images) and augmented these by considering all convex patch combinations. [13] proposed a bi-directional image synthesis approach for non-rigid registration of the pelvic area, where random forests are trained on Haar-like features extracted from

pairs of pre-aligned CT and MR patches. An auto-context model was used to incorporate neighboring prediction results. All these learning-based approaches depend on co-registered multi-modal images for training. This is very difficult for deforming structures as dense (voxel-wise) spatial correspondences are required and CT and MR images cannot be acquired simultaneously yet [14].

Learning without paired data. A cross-modality synthesis method which does not require paired data was proposed in [15]. It is based on generating multiple target modality candidate values for each source voxel independently using cross-modal nearest neighbor search. A global solution is then found by simultaneously maximizing global MI and local spatial consistency. Finally, a coupled sparse representation was used to further refine the synthesized images. When applied to T1/T2 brain MRIs, T1 images were better synthesized than T2 images (0.93 vs. 0.85 correlation to ground truth). Extending the method to a supervised setting outperformed state-of-the-art supervised methods slightly.

Recently cyclic-consistent Generative Adversarial Networks (cycle-GANs) were proposed for learning an image-to-image mapping between two domains (\mathcal{A} & \mathcal{B}) from unpaired datasets [16]. The method is based on two generator networks (G_B to synthesize image $\hat{\mathbf{I}}_B$ from \mathbf{I}_A , G_A) and two discriminator networks (D_A , D_B). Besides the usual discriminator loss to differentiate synthesized and real images (e.g. $\hat{\mathbf{I}}_A$, \mathbf{I}_A), a cycle loss was introduced which measures the difference between the real image and its twice synthesized image, e.g. $|\mathbf{I}_A - G_A(G_B(\mathbf{I}_A))|_1$. Good performances were shown for various domain translation tasks like labels to photos and aerial photos to maps. Very recently, this promising approach was employed for slice-wise synthesizing CT from MR head images from unpaired data [2]. It achieved lower mean squared errors (74 vs. 89 HU) than when training the same generator network on rigidly aligned MR and CT data [17]. It was reasoned that this could be due to misalignments, as the images contained also deforming structures (e.g. neck, mouth). Cycle-GANs were used for synthesis of MR from unpaired CT images for enriching a cardiac dataset for training thereafter a segmentation network [18]. A view alignment step using the segmentations was incorporated to make the layout (e.g. position, size of anatomy) of the CT and MR images similar, such that the discriminator cannot use the layout to differentiate between them. Furthermore the myocardium mask for both modalities was provided during training, as the cycle-GAN not only changed the intensities but also anatomical locations such that the mask was no longer in correspondence with the image. Hence this is not a completely unsupervised approach. Similarly, a shape-consistency loss from segmentations was incorporated in [19] to avoid geometric changes between the real and synthesized images. It was argued that "from the discriminator perspective, geometric transformations do not change the realness of synthesized images since the shape of training data is arbitrary". However this does not hold if there is a geometric bias between the two datasets.

Synthesized MR PD/T1 brain images via patch matching were shown to be useful for segmentation and inter-modality cross-subject registration [20]. If this also holds for MR-CT synthesis via cycle-GANs for thoracic and abdominal

regions has not yet been studied. Our contributions include (i) combining two state-of-the-art multi-modal DIR similarity measures (NMI, MIND), (ii) studying the effect of the image region size on the consistency of the synthesized 3D images, and (iii) evaluating the usefulness of synthesized images for deformable registration of CT and MR images from the thorax and abdomen against a strong baseline.

2 Materials

We used 17 unpaired and 3 paired 3D MR-CT images from the VISCERAL Anatomy3 benchmark training set (unenhanced, whole body, MR-T1) and their gold standard segmentations for evaluation [21]. The 3 subjects with paired data had IDs 10000021, 10000067 and 10000080. All MRIs were bias field corrected using the N4ITK method [22]. All images were resampled to an isotropic resolution of 1.25 mm. This was motivated by the image resolution of the original MRIs being $1.25 \times 6 \times 1.25$ mm³ in left-right, posterior-anterior and superior-inferior direction. The CT images had a resolution between $0.8 \times 0.8 \times 1.5$ mm³ and $1.0 \times 1.0 \times 1.5$ mm³.

To reduce memory requirements, we automatically extracted from each image two regions such that the included gold standard segmentations were at least 5 mm away from the inferior and superior region boundary. The thorax region covered the segmentations of the liver, spleen, gallbladder, and right and left lung. The abdominal region contained the bladder, lumbar vertebra 1, right and left kidney, and right and left psoas major muscle, see Figs. 2, 1, left column.

Closer investigation of poor performing registration results showed that for case 10000067 the segmentation labels of the right and left kidney were swapped in the MRI. Additionally, for 10000080 the segmentation of the lumbar vertebra 1 in the MRI seems to be that of lumbar vertebra 2. We corrected these kidney annotations and excluded this lumbar vertebra 1 segmentations from the results.

3 Method

3.1 Image Synthesis

Cycle-GAN. For image synthesis, we followed the cycle-GAN network architecture as described in [16,2], starting from an existing implementation¹. In short, the two generators (G_{CT} , G_{MR}) are 2D fully convolutional networks with 9 residual blocks and two fractionally strided convolution layers (res-net). The discriminators (D_{CT} , D_{MR}) are fully convolutional architectures to classify overlapping $P \times P$ image patches as real or fake (PatchGAN) [23]².

¹ <https://github.com/xhujoy/CycleGAN-tensorflow>

² The discriminators consist of 5 convolutions layers (I256-C128-C64-C32-C32, stride length 2-2-2-1-1, 4×4 kernels) for $P=70$ and 4 layers (I256-C128-C64-C64, stride length 2-2-1-1) for $P=34$. Leaky ReLU activation functions (factor 0.2) were used. Data was normalized by instance normalization.

The networks take input images of size 256×256 pixels and C channels. Larger-sized test images were synthesized from the average result of $256 \times 256 \times C$ regions extracted with a stride length of $S \times S \times S_C$. The cycle-GAN was optimized to reduce the overall loss L , which is a weighted sum of the discriminator losses L_{CT} , L_{MR} and the generator cyclic loss L_{cyc} :

$$L = L_{\text{CT}} + L_{\text{MR}} + \lambda_{\text{cyc}} L_{\text{cyc}} \quad (1)$$

$$L_{\text{CT}} = (1 - D_{\text{CT}}(I_{\text{CT}}))^2 + D_{\text{CT}}(G_{\text{CT}}(I_{\text{MR}}))^2 \quad (2)$$

$$L_{\text{MR}} = (1 - D_{\text{MR}}(I_{\text{MR}}))^2 + D_{\text{MR}}(G_{\text{MR}}(I_{\text{CT}}))^2 \quad (3)$$

$$L_{\text{cyc}} = \|G_{\text{CT}}(G_{\text{MR}}(I_{\text{CT}})) - I_{\text{CT}}\|_1 + \|G_{\text{MR}}(G_{\text{CT}}(I_{\text{MR}})) - I_{\text{MR}}\|_1 \quad (4)$$

3.2 Image Registration

Rigid Registration. The CT and MR images were first rigidly registered using the function *imregister* from the MATLAB Image Processing Toolbox [24], set to multi-modal configuration (Mattes Mutual Information, one-plus-one evolutionary optimization). These rigid registration results were then used as starting points for all subsequent deformable image registrations.

Deformable Registration - MIND. The so-called modality independent neighborhood descriptor (MIND) was proposed as dissimilarity measure for multi-modal DIR [7]. MIND is based on a multi-dimensional descriptor \mathbf{s}_{MIND} per voxel \mathbf{x} , which captures the self-similarity of the image patch around \mathbf{x} (denoted as $\mathbf{P}(\mathbf{x})$) with the patches $\mathbf{P}(\mathbf{x} + \mathbf{r})$ in a local neighborhood \mathcal{N} of \mathbf{x} . The single entries $\mathbf{s}_{\text{MIND}}(\mathbf{I}, \mathbf{x}, \mathbf{r}_i)$ are calculated by a Gaussian function

$$\mathbf{s}_{\text{MIND}}(\mathbf{I}, \mathbf{x}, \mathbf{r}_i) = \frac{1}{n} \exp \left(-\frac{d_p(\mathbf{I}, \mathbf{x}, \mathbf{r}_i)}{v(\mathbf{I}, \mathbf{x})} \right) \quad (5)$$

where n is a normalization constant such that the maximum value in \mathbf{s}_{MIND} is 1, d_p defines the patch dissimilarity $d_p(\mathbf{I}, \mathbf{x}, \mathbf{r}) = \sum_{\mathbf{x}_j \in \mathbf{P}(\mathbf{x})} \mathbf{G}_\sigma(\mathbf{x}_j)(\mathbf{P}(\mathbf{x}_j) - \mathbf{P}(\mathbf{x}_j + \mathbf{r}))^2$ with Gaussian kernel \mathbf{G}_σ of the same size as patch $\mathbf{P}(\mathbf{x})$ and the half-size of the patch being equal to $[1.5\sigma]$. v is the variance of a six-neighborhood search region. \mathbf{s}_{MIND} is calculated in a dense fashion for each image independently. The dissimilarity $E_{\text{MIND}}(\mathbf{A}, \mathbf{B})$ of images \mathbf{A} and \mathbf{B} is finally defined by $E_{\text{MIND}}(\mathbf{A}, \mathbf{B}) = \sum_{\mathbf{x} \in \Omega} E_{\text{MIND}}(\mathbf{A}, \mathbf{B}, \mathbf{x})^2$ with

$$E_{\text{MIND}}(\mathbf{A}, \mathbf{B}, \mathbf{x}) = \frac{1}{|\mathcal{N}|} \sum_{\mathbf{r}_i \in \mathcal{N}} |\mathbf{s}_{\text{MIND}}(\mathbf{A}, \mathbf{x}, \mathbf{r}_i) - \mathbf{s}_{\text{MIND}}(\mathbf{B}, \mathbf{x}, \mathbf{r}_i)|. \quad (6)$$

In the MIND registration framework, the images are downsampled via Gaussian pyramids and the deformation field is regularized via the squared L2-norm. Additionally after each Gauss-Newton update step during optimization, each deformation field is replaced by combining half of its own transformation with half of the inverse transformation of the other deformation field (see [7,25]) to obtain diffeomorphic transformations. We used the provided code [26] and compared results after integrating the MIND measure in our DIR method [27].

Deformable Registration - ourDIR. We extended our DIR method, based a linearly interpolated grid of control points and various displacement regularization measures, to incorporate the multi-modal (dis)similarity measures normalized mutual information (NMI) and MIND, and their combination NMI+MIND.

Given fixed image \mathbf{I}_f , moving image \mathbf{I}_m , displacements \mathbf{k} at the control points, and interpolation function d to get dense displacements, the NMI dissimilarity $E_{\text{NMI}}(\mathbf{I}_f, \mathbf{I}_m(d(\mathbf{k})))$ is defined by $E_{\text{NMI}}(\mathbf{A}, \mathbf{B}) = -(H_A + H_B)/H_{\mathbf{A}, \mathbf{B}}$, with marginal entropies H_A, H_B and joint entropy $H_{\mathbf{A}, \mathbf{B}}$ computed from intensity histograms with 100 equally-spaced bins between the 0.5 and 99.5 percentiles of the image intensity. The gradients of $E_{\text{NMI}}(\mathbf{I}_f, \mathbf{I}_m(d(\mathbf{k})))$ with respect to $d(\mathbf{k})^{(i)}[x]$ are calculated as described in [28]. To avoid infinity gradients, we replace zero probabilities with $1/(2N_V)$, where N_V is the number of image voxels. We combined the dissimilarities NMI and MIND by

$$E_{\text{N+M}}(\mathbf{I}_f, \mathbf{I}_m(d(\mathbf{k}))) = \beta E_{\text{NMI}}(\mathbf{I}_f, \mathbf{I}_m(d(\mathbf{k}))) + (1 - \beta) s E_{\text{MIND}}(\mathbf{I}_f, \mathbf{I}_m(d(\mathbf{k}))) \quad (7)$$

where s is a scaling parameter to get E_{MIND} in the same range [29] and $\beta \in [0, 1]$ is a weighting term. The choice of s is not trivial, as the magnitude of change per dissimilarity measure from initial to ideal knot displacements $D_{\text{init,ideal}}(E_{\text{dissim}})$ is unknown. We tested 3 strategies, namely (i) using a fixed parameter s , (ii) using the initial gradient magnitude via

$$s = \frac{D_{\text{init,ideal}}(E_{\text{NMI}})}{D_{\text{init,ideal}}(E_{\text{MIND}})} \approx \frac{\|\nabla E_{\text{NMI}}(\mathbf{I}_f, \mathbf{I}_m(d(\mathbf{k}_{\text{init,q}})))\|_2}{\|\nabla E_{\text{MIND}}(\mathbf{I}_f, \mathbf{I}_m(d(\mathbf{k}_{\text{init,q}})))\|_2} \quad (8)$$

or (iii) basing it on the change in dissimilarity during registration:

$$s = \frac{D_{\text{init,ideal}}(E_{\text{NMI}})}{D_{\text{init,ideal}}(E_{\text{MIND}})} \approx \frac{|E_{\text{NMI}}(\mathbf{I}_f, \mathbf{I}_m) - E_{\text{NMI}}(\mathbf{I}_f, \mathbf{I}_m(d(\mathbf{k}_{\text{init,q}})))|}{|E_{\text{MIND}}(\mathbf{I}_f, \mathbf{I}_m) - E_{\text{MIND}}(\mathbf{I}_f, \mathbf{I}_m(d(\mathbf{k}_{\text{init,q}})))|}. \quad (9)$$

The final cost function is $F(d(\mathbf{k})) = E_{\text{dissim}}(\mathbf{I}_f, \mathbf{I}_m(d(\mathbf{k}))) + \lambda R(\mathbf{k})$ where $R(\mathbf{k})$ regularizes the displacements at the control points by their TV or L2 norm.

4 Experiments and Results

4.1 Image Synthesis

Images intensities were linearly scaled to [0,255]. Suitable image regions were cropped to fit the network size instead of resizing the image in-plane [2], as resizing lead to distortions in the synthesized images due to systematic differences in image size between the two modalities. Image regions (ROIs) of size $286 \times 286 \times C$, for $C \in \{3, 12\}$ were extracted randomly from the training data. Dark ROIs, with a mean intensity of less than 1/4 of that of the ROI in the center of the 3D image, were not selected to avoid including a lot of background. ROIs were further randomly cropped to $256 \times 256 \times C$ during network training. A Cycle-GANs was trained for 200 epochs per image region (thorax or abdomen). We used a training regime as previously reported [2], namely Adam optimizer, learning rate fixed to 0.0002 for 1-100 epochs and linearly reduced to 0 for 101-200 epochs, $\lambda_{\text{cyc}}=10$. 3D test images were created using an in-plane stride length of $S=4$ and a channel stride length of $S_C=2$ for $C=3$ and $S_C=4$ for $C=12$.

4.2 Image Registration

MR-CT deformable image registration based on image multi-modal similarity measure MIND, NMI, or NMI+MIND was compared with image synthesis and then deformable registration using local NCC as image similarity. All registrations used ourDIR framework with total variation regularization. Registration parameters were optimized via grid search. These are the weighting of regularization term $\lambda \in \{0.0125, 0.025, 0.05, 0.1, 0.2\}$, the control point spacing $s \in \{8, 10, 12, 14, 16\}$ pixels, and the number of multi-resolution levels $l \in \{2, 3, 4\}$. The best strategy for combining NMI and MIND was using the initial gradient magnitude, i.e. Eq. (8), and $\beta=0.8$.

4.3 Results

Example of synthesized images are shown in Figs. 1-2. Inconsistency across slices can be seen when ROIs with few slices ($C=3$) are used. Synthesized image structures do mostly not adhere to the contours of the lung segmentations from the real image. To be realistic, the generators had to learned the substantial bias in lung volume between the two modalities, see Table 1. CT images were generally acquired in end-inhale state, while MRIs in end-exhale. Changing the patch-GAN discriminator to a shallower architecture, such that each output node has a receptor field of $P=34 \times 34$ instead of 70×70 could sometimes reduced the misalignment of the lung segmentations (e.g. Figs. 1d), but was less powerful in image synthesis (e.g. region between the lungs in Fig. 2d).

The performance of the deformable image registration for the original images (CT, MR) and the cycle-GAN synthesized images are listed in Table 2. While synthesized images can achieve a similar overlap than multi-modal NMI registration for the abdomen (77.4 vs. 78.8%), they are substantially worse for the thorax due to the bias in lung volume. Performance of synthesized CT images, which were more affected by synthesized volume lung volume changes, was generally lower than for synthesized MRIs. Results are shown in Fig. 3.

	Liver	Spleen	Gallb.	rLung	lLung	Bladder	lVert1	rKidney	lKidney	rPsoas	lPsoas
meanCT	1896	244	42	2598	2253	198	63	185	197	208	188
meanMR	1576	248	201	1338	1144	153	62	211	225	158	180
Ratio (%)	120	98	21	194	197	129	101	88	88	131	105

Table 1: Mean volume of segmented structures in cm^3 per modality and ratio meanCT/meanMR for unpaired training data.

	CT-MR				Synthesized MR				Synthesized CT			
	NMI+		$C=3$	$C=12$	$C=12$	$C=12$	$P=70$		$P=70$		$P=70$	
	rigid	MIND					NMI	MIND	$P=34$	$P=22$	$P=34$	$P=22$
Thorax	55.2	65.4	75.2	75.7	65.2	62.2	62.3	55.4	58.1	59.3	54.6	55.2
Abdomen	60.6	73.9	78.8	78.3	66.6	76.9	77.4	58.0	48.4	72.3	72.6	60.4
Both	59.6	67.7	76.7	76.9	65.6	69.0	69.1	56.6	52.5	64.6	63.1	57.8

Table 2: DIR performance measured by mean Dice overlap ratio (%) for original images (CT-MR) or for cycle-GAN synthesized MR or CT images, using ROIs of size $256 \times 256 \times C$ with discriminator based on $P \times P$ patches. Results within 5% to the best result are marked in **bold**.

5 Conclusion

We combined two multi-modal image similarity measure (NMI, MIND) and observed a similar performance as when using only NMI, and in contrast to [7] no improvement of MIND over NMI.

We investigated the usefulness of a fully unsupervised MR-CT image modality synthesis method for deformable image registration of MR and CT images. Against the established multi-modal deformable registration methods, synthesizing images via cycle-GAN and then using a robust mono-modal image similarity measure achieved at best a similar performance. In particular one has to be careful to have collections of the two image modalities which are balanced, i.e. not biased for a modality, as such differences are readily synthesized by the cycle-GAN framework. Ensuring that synthesized images are truly in spatial correspondence with the source image would require incorporating a deformable image registration into the cycle-GAN.

Acknowledgments: We thank the EUs 7th Framework Program (Agreement No. 611889, TRANS-FUSIMO) for funding and acknowledge NVIDIA for GPU support.

References

1. Sotiras, A., Davatzikos, C., Paragios, N.: Deformable medical image registration: A survey. *IEEE Trans Med Imag* **32(7)** (2013) 1153–1190
2. Wolterink, J., Dinkla, A., Savenije, M., Seevinck, P., van den Berg, C., Isgum, I.: Deep mr to ct synthesis using unpaired data. In: Int Workshop on Simulation and Synthesis in Medical Imaging, Springer (2017) 14–23
3. Studholme, C., Hill, D., Hawkes, D.: An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition* **32(1)** (1999(32)) 71–86
4. Pluim, J., Maintz, J., Viergever, M.: Image registration by maximization of combined mutual information and gradient information. In: Medical Image Computing and Computer-Assisted Intervention, Springer (2000) 452–461
5. Haber, E., Modersitzki, J.: Intensity gradient based registration and fusion of multi-modal images. *Medical Image Computing and Computer-Assisted Intervention* (2006) 726–733

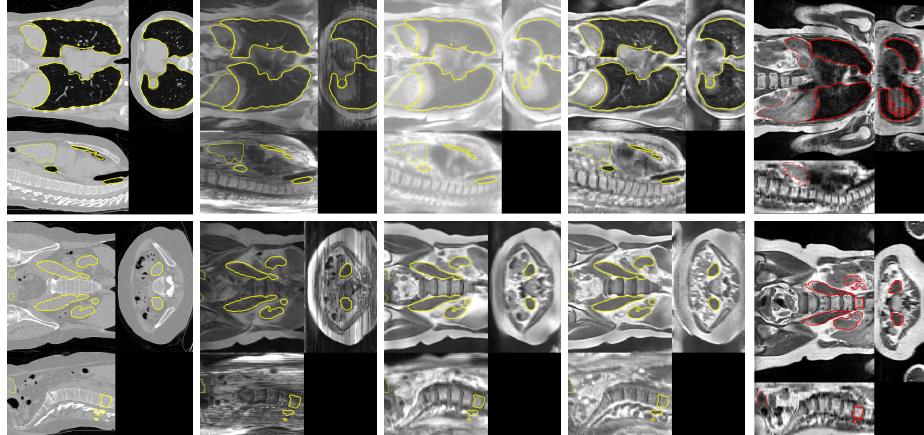


Fig. 1: Illustration of MR synthesis from CT for (top) thoracic and (bottom) abdominal region. (a) original CT, (b-d) synthesized MRIs from (b) $256^2 \times 3$ ROIs, (c) $256^2 \times 12$ ROIs, (d) $256^2 \times 12$ ROIs and 34×34 patches, (e) original MRI. Original MR (CT) contours in red (yellow).

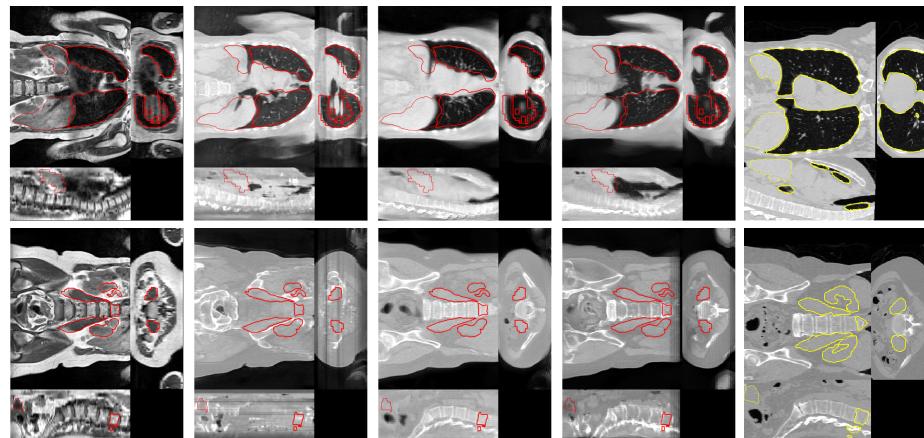


Fig. 2: Illustration of CT synthesis from MR for (top) thoracic and (bottom) abdominal region. (a) original MRI, (b-d) synthesized CTs from (b) $256^2 \times 3$ ROIs, (c) $256^2 \times 12$ ROIs, (d) $256^2 \times 12$ ROIs and 34×34 patches, (e) original CT rigidly aligned. Original MR (CT) contours in red (yellow).

6. Wachinger, C., Navab, N.: Entropy and Laplacian images: Structural representations for multi-modal registration. *Med Image Anal* **16**(1) (2012) 1–17
7. Heinrich, M., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F., Brady, M., Schnabel, J.: MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration. *Med Image Anal* **16**(7) (2012) 1423–1435

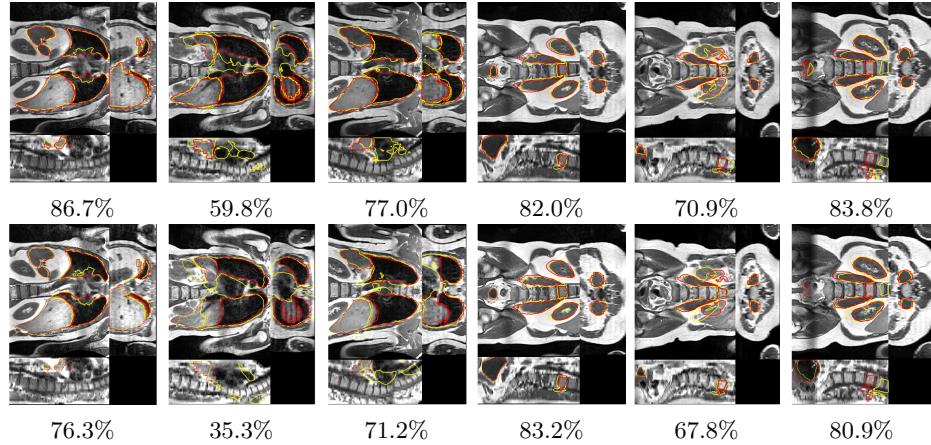


Fig. 3: Deformable image registration results for (top) CT-to-MRI based on NMI, (bottom) synthesizedMRI-to-MRI based on local NCC ($256^2 \times 12$ ROIs, 34×34 patches). Image: original MR, yellow contour: original MR, red contour: deformed CT or synthesized MRI, value: mean Dice.

8. Andronache, A., von Siebenthal, M., Székely, G., Cattin, P.: Non-rigid registration of multi-modal images using both mutual information and cross-correlation. *Med Image Anal* **12**(1) (2008) 3–15
9. Knops, Z., Maintz, J., Viergever, M., Pluim, J.: Registration using segment intensity remapping and mutual information. In: *Medical Image Computing and Computer-Assisted Intervention*, Springer (2004) 805–812
10. Cheng, X., Zhang, L., Zheng, Y.: Deep similarity learning for multimodal medical images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* (2016) 1–5
11. So, R., Chung, A.: A novel learning-based dissimilarity metric for rigid and non-rigid medical image registration by using Bhattacharyya distances. *Pattern Recognition* **62** (2017) 161–174
12. Roy, S., Carass, A., Jog, A., Prince, J., Lee, J.: MR to CT registration of brains using image synthesis. In: *Proc SPIE*. Volume 9034., NIH Public Access (2014)
13. Cao, X., Yang, J., Gao, Y., Guo, Y., Wu, G., Shen, D.: Dual-core steered non-rigid registration for multi-modal images via bi-directional image synthesis. *Med Image Anal* (2017)
14. Liu, Y.H., Sinusas, A.: *Hybrid Imaging in Cardiovascular Medicine*. CRC Press (2017)
15. Vemulapalli, R., Van Nguyen, H., Kevin Zhou, S.: Unsupervised cross-modal synthesis of subject-specific scans. In: *ICCV*. (2015) 630–638
16. Zhu, J.Y., Park, T., Isola, P., Efros, A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv:1703.10593* (2017)
17. Nie, D., Trullo, R., Lian, J., Petitjean, C., Ruan, S., Wang, Q., Shen, D.: Medical image synthesis with context-aware generative adversarial networks. In: *Medical Image Computing and Computer-Assisted Intervention*, Springer (2017) 417–425

18. Chartsias, A., Joyce, T., Dharmakumar, R., Tsafaris, S.: Adversarial image synthesis for unpaired multi-modal cardiac data. In: Int Workshop on Simulation and Synthesis in Medical Imaging, Springer (2017) 3–13
19. Zhang, Z., Yang, L., Zheng, Y.: Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2018) 9242–9251
20. Iglesias, J., Konukoglu, E., Zikic, D., Glocker, B., Van Leemput, K., Fischl, B.: Is synthesizing MRI contrast useful for inter-modality analysis? In: Medical Image Computing and Computer-Assisted Intervention, Springer (2013) 631–638
21. Jimenez-del Toro, O., Müller, H., Krenn, M., Gruenberg, K., Taha, A., Winterstein, M., Eggel, I., Foncubierta-Rodríguez, A., Goksel, O., Jakab, A., et al.: Cloud-based evaluation of anatomical structure segmentation and landmark detection algorithms: VISCERAL anatomy benchmarks. *IEEE Trans Med Imag* **35(11)** (2016) 2459–2475
22. Tustison, N., Avants, B., Cook, P., Zheng, Y., Egan, A., Yushkevich, P., Gee, J.: N4ITK: improved N3 bias correction. *IEEE Trans Med Imag* **29(6)** (2010) 1310–1320
23. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.: Image-to-image translation with conditional adversarial networks. *arXiv:1611.07004* (2016)
24. MathWorks: Matlab imregister documentation. <https://ch.mathworks.com/help/images/ref/imregister.html> (2017) Online. Accessed: 04-December-2017.
25. Avants, B., Epstein, C., Grossman, M., Gee, J.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal* **12(1)** (2008) 26–41
26. Heinrich, M.: Symmetric Gauss-Newton deformable registration code. <http://www.ibme.ox.ac.uk/research/biomedia/julia-schnabel/files/symgn.zip> (2012) note = "[Online; accessed 26-November-2017]".
27. Vishnevskiy, V., Gass, T., Szekely, G., Tanner, C., Goksel, O.: Isotropic total variation regularization of displacements in parametric image registration. *IEEE Trans Med Imag* **36(2)** (2017) 385–395
28. Crum, W., Hill, D., Hawkes, D.: Information theoretic similarity measures in non-rigid registration. In: Information Processing in Medical Imaging, Springer (2003) 378–387
29. Lundqvist, R., Bengtsson, E., Thurfjell, L.: A combined intensity and gradient-based similarity criterion for interindividual SPECT brain scan registration. *EURASIP Journal on Advances in Signal Processing* **2003(5)** (2003) 967364